DOCUMENT RESUME

ED 057 410 24 CG 006 966

AUTHOR	Gorsuch, Richard L.
TITLE	Value Conflict in the School Setting. Final Report.
INSTITUTION	George Peabody Coll. for Teachers, Nashville, Tenn. Inst. on Mental Retardation and Intellectual Development.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO	BR-9-0427
PUB DATE	31 Aug 71
GRANT	OEG-4-9-520427-0047-010
NOTE	227p.
EDRS PRICE	MF-$0.65 HC-$9.87
DESCRIPTORS	*Academic Achievement; Behavior Problems; Classroom Environment; *Conflict; Personality Problems; Student Attitudes; *Student Subcultures; *Teacher Influence; *Values

ABSTRACT

This project examined how values relate to teachers' ratings of their children and teacher reports of critical incidents which happened in the classrooms. The role of race and social class differences in values and classroom problems was also investigated. An examination of children's remarks showed that Good Manners, Kindness, Academic Achievement, Diligence, and Obedience were among the more frequently used value categories. Analysis of the use of the categories and of responses to further value questionnaires developed to be effective within each of the subcultures showed only minor differences in values as a function of social class, sex and race. Most of the classroom incidents were judged to involve values. Value differences between the child and the teacher or the child and his peers were predictive of involvement in incidents and of the teacher's rating the child as having personality or behavior problems. Values also increased the prediction of grades over that which could be achieved by using IQ alone. Teachers did have impact on the children's values. Children shifted towards their teachers over the semester if the teacher scored higher than the children on the value scales. (Author)

ED057410

9-0427
PA 24

Final Report

Project No. 9-0427
Grant No. OEG-4-9-520427-0047-010

Richard L. Gorsuch
The John F. Kennedy Center for Research
on Education and Human Development
George Peabody College for Teachers
Nashville, Tennessee 37203

VALUE CONFLICT IN THE SCHOOL SETTING

August 1971

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

National Center for Educational Research and Development
(Regional Research Program)

CG 006 966

ED057410

Final Report

Project No. 9-0427

Grant No. OEG-4-9-520427-0047-010

VALUE CONFLICT IN THE SCHOOL SETTING

Richard L. Gorsuch

George Peabody College for Teachers

Nashville, Tennessee

August 31, 1971

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EEUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

Final Report

Project No. 9-0427

Grant No. OEG-[illegible]-9-520427-0047-010

VALUE CONFLICT IN THE SCHOOL SETTING

Richard L. Gorsuch

George Peabody College for Teachers

Nashville, Tennessee

August 31, 1971

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EEUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	ACKNOWLEDGEMENTS	viii
I	NATURE AND SCOPE OF THE STUDY	1
	The Problems	1
	Subcultural Differences in Values	1
	Values and Behavior	3
	The Impact of Value Discrepancies	5
	Prerequisite Questions	7
	The Nature of Children's Values	7
	Value Shifts	8
	Teacher Perceptions	9
	The Interrelatedness of the Objectives	9
II	RESEARCH STRATEGY: A FIELD STUDY	11
	Studying Values	11
	Subject Samples	12
	Design	12
	Studies I, II and III	13
	Bias Analysis	14
	Procedures	17
	Data Collection	17
	Measuring SES	19
	Testing Materials	21
	Missing Data Procedures	24
	Statistical Analyses	24
	Design Limitations	26
III	CHILDREN'S PHENOMENOLOGICAL VALUE CATEGORIES	28
	On Developing A Set of Value Categories	28
	Pilot Study	29
	Development of the Free Response Form	29
	Preliminary Value Categories	30
	Study I: Collecting and Categorizing Value Statements	35
	Ethnic Differences in Value Categories	38
IV	MEASUREMENT OF CHILDREN'S VALUES	43
	Measuring Ethnic Differences Fairly	43
	Types of Scales	44
	Item Development and Administration	46
	Scott Items	46
	Osgood Items	47
	Administration: Study II	47
	Similarity of Scott and Osgood Scales	47
	Correlations Between A Priori Scales	48
	Item Factor Analyses	51
	The Resulting Scales	54
	Characteristics of the Scales in Study III	57
	Content and Meaning of the Scales	62

TABLE OF CONTENTS (Continued)

CHAPTER	TITLE	PAGE
V	VALUE DISCREPANCIES AS A FUNCTION OF SOCIAL CLASS, SEX AND RACE	64
	The Relationship of the Teachers' Values to the SES and Racial Composition of their schools	64
	Social Class, Sex and Race Differences in Children's Values.	67
	The State of the Literature	67
	Problems in Analyzing Ethnic Differences .	68
	Group Differences in Values: Study II . .	69
	Group Differences in Study III	70
	The Importance of Differences in Group Values for the School Setting	74
VI	SHIFTS IN VALUES AND VALUE DISCREPANCIES OVER A SEMESTER	78
	Stability and Change in Values	78
	Changes in Pupil-Teacher Value Discrepancies.	80
	The Analyses	80
	Time Trends	80
	Group Interactions with Time	82
	Individual Values.	86
	Changes in Pupil-Peer Value Discrepancies Over the Semester.	87
	A Tantalizing Question: Sources of Change .	87
VII	TEACHER PERCEPTIONS AND LABELING: AN APPROACH TO CONFLICT MANIFESTATIONS.	91
	The Development of Teacher Perceptions: A Model.	91
	The Measurement of Teacher Perceptions. .	93
	The Structure of Perception: Factor Analysis of Teacher Ratings	95
	Perceptions of Pupils Grouped by Social Class, Sex, and Race.	96
	Social Class Differences	96
	Sex Differences	99
	Race Differences	99
	Perceptual Shifts: A Time Perspective. . .	101
	Teacher Labeling Style and Important Incidents	103
	Discussion	106
	Structural Analyses	107
	Group Differences.	107
	Perceptual Shifts.	107
	Teacher Labeling Style and Important Incidents	108

TABLE OF CONTENTS (Continued)

CHAPTER	TITLE	PAGE
VIII	VALUE CONFLICTS AND OTHER OUTCOMES	109
	Values and Classroom Incidents	109
	The Occurrence of Value Conflicts	109
	Group Differences in Value Conflicts	109
	Value Conflicts as a Function of Values and Value Discrepancies	111
	Other Outcomes as a Function of Values and Value Discrepancies	115
	Anxiety	115
	Grades	116
	The Prediction of Teacher Perceptions	119
	Personality Problem Ratings	119
	Behavior Problem Ratings	124
	Values and Value Discrepancies as Predictive Constructs	126
IX	SUMMARY AND CONCLUSIONS	129
	The Research Project	129
	Teachers and Values	130
	The Impact of Values	130
	Value Conflicts, Value Discrepancies and Subcultures	131
REFERENCES		132
APPENDICES		
A	Race of Administrator Effect	140
B	Parental Permission Letter	144
C	The Accuracy of Teachers' Ratings of Students' Social Class	146
D	Value Measurements Developed by the Project	176
E	Special Measuring Instruments Used by the Project	202
F	Dangers In Using The Children's Locus of Control Scale With Children	236
G	A Comparison of Several Techniques for Measuring Children's Values	268
H	Sequential Multiple Regression Analyses	274

LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	General Information on Samples Used	13
2.2	Sex, Race and SES Composition of Samples	14
2.3	Means of the Participants and Non-Participants on Available Data	17
2.4	School Reputation and Teachers' Estimation of SES	20
2.5	Basic Measurements: Study III	22
2.6	Reliabilities: Study III	24
3.1	List of Initial Value Categories Used in Addition to Scott's	32
3.2	Rater's Reliabilities	33
3.3	Children's Value Categories	34
3.4	Value Categories: Their Frequency of Use and Reliability	37
3.5	Group Means for Total Responses in Study I	39
3.6	Value Category Usage by Sex, Race and SES of Pupil	40
4.1	Correlations Between Scott and Osgood Items	48
4.2	Reliabilities of and Correlations Among Scott A Priori Scales: Lower-Class Blacks	50
4.3	Reliabilities of Adjective Scales and Correlations Among Semantic Differential Measures of Values: Middle-Class Whites	51
4.4	Factor Loadings of Scott Items (.40 and Above) on Two Randomly Divided Halves of Questionnaire	53
4.5	Value Socialization Items	55
4.6	Semantic Differential Adjectives	56
4.7	Value Concepts for Semantic Differential Rating and Rank Ordering	57

LIST OF TABLES (Continued)

TABLE NO.	TITLE	PAGE
4.8	Reliabilities of the Value Scales within SES x sex x race subgroups	58
4.9	Correlations Among Value Scales	60
4.10	Correlations of Value Scales with Personality Characteristics	61
5.1	Summary of Analyses of Teachers' Values by School Race and School SES	65
6.1	Teachers' Values	79
7.1	Teacher Perception and Labeling of Behavior as an Acquaintanceship Process	94
7.2	Results of Factor Analysis of Teacher Ratings	96
7.3	Self-Anchoring Scale Ratings: Behavior Problem x Personality Problem x Time Interaction	103
8.1	Beta Weights for Predicting Value Conflicts	114
8.2	Beta Weights for Predicting Mid-Year Anxiety	116
8.3	Beta Weights for Predicting Grade Point Average	117
8.4	Beta Weights for Predicting Personality Problems	120
8.5	Beta Weights for Predicting Behavior Problems	124

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
2.1	Sampling design	12
2.2	Separation of the subjects and non-participants in two dimensional space	16
3.1	Usage of value categories by extreme groups	42
5.1	Semantic Differential Total Value means for sex, race, SES, and grade level groups	72
5.2	Value Socialization means for SES, sex, race, and grade level groups	73
6.1	Change in teacher-child discrepancies over time and grade: Value Socialization Scale	81
6.2	Change in teacher-child discrepancies over time and grade: Semantic Differential Total Value Score	82
6.3	Sex and SES differences in change over time: Teacher-pupil discrepancies on the Semantic Differential Total Value Score	83
6.4	Racial differences in change over time: Teacher-pupil discrepancies on Semantic Differential Total Value Score	84
6.5	Racial differences in change over time: Pupil's Semantic Differential Total Value Score	85
6.6	Racial differences in change over time: Subjects scoring above teacher only	86
6.7	Change in Value Socialization over time as a function of teacher and peer discrepancy	89
7.1	Self-Anchoring Scale Ratings: Class effect	97
7.2	Behavior Problem Ratings: Class x race interactions	98
7.3	Personality Problem Ratings: Class x race interaction	98
7.4	Important Incident Frequency: Class x type of incident interaction	99

019

LIST OF FIGURES (Continued)

FIGURE NO.	TITLE	PAGE
7.5	Self-Anchoring Scale Ratings: Race x sex interaction	100
7.6	Behavior Problem Ratings: Sex effect	100
7.7	Important Incident Frequency: Sex x type incident interaction	101
7.8	Self-Anchoring Scale Ratings: Race x time interaction	102
7.9	Self-Anchoring Scale Ratings: Important Incident type x time interaction	104
7.10	Behavior Problem Ratings: Important Incident type x time interaction	104
7.11	Self-Anchoring Scale Ratings: Important Incident type x teacher rating style interaction	105
7.12	Behavior Problem Ratings: Important Incident type x teacher-pupil discrepancy interaction	106
8.1	Value-related and non-value related incidents by social class and sex	110
8.2	Percent of value incidents by sex and social class (school)	112
8.3	Predicting grades from teacher-pupil discrepancy on the Semantic Differential Total Value Scores	117
8.4	Predicting Personality Problem Ratings from discrepancies on the Semantic Differential Total Value Scores	121
8.5	Relationship of teacher-pupil Value Socialization discrepancies to Personality Problem Rating	122
8.6	Race differences in predicting Behavior Problems from teacher-pupil Value Socialization discrepancies	125

ACKNOWLEDGEMENTS

We would like to thank the Nashville-Davidson County Metropolitan Public School System for making this research possible. Dr. Edward Binkley of Metro Schools was instrumental in providing the necessary entry to the schools and in advising on suitable materials and procedures. The warm cooperation of the sixteen schools, the fifty-three classrooms and the over 1100 children was particularly gratifying. We hope that their faith in the importance of research was justified by this and other such projects.

Dr. J. Robert Newbrough brought to the project his transactional perspective and support, both of which we deeply appreciated. He and Donald Bartlett pursued the line of thought which resulted in Chapter 7.

Ruth Smith and Richard Henighan shared in the general administration of the project. This involved supervising the collection of data and the carrying out of many of the analyses. The former was deeply involved at all stages of the project but particularly so in the materials for Chapters 3 and 4 and in the preparation of the final report. The latter was particularly involved in Chapter 6. Vital aid and support were also provided by Alice Dawson and, especially, Betty Howard.

The contributions of graduate assistants were instrumental in providing the sophisticated aid needed at each step of the project. David Tinsley and Malcolm Scherz helped develop the rating categories that formed the basis of Chapter 3. William Wallace also participated in the application of the categories for Chapter 3 and in the development of the measuring instruments (Chapter 4). Phil Osborne not only aided in the data collection and analysis of Study III but also rough drafted Chapter 5. Morris Beers and Lloyd Richardson performed the extensive -- and tiring -- analyses for Chapter 8. The contributions of Vera Stevens were also greatly appreciated.

The permission of the Institute for Personality and Ability Testing to reproduce scale B of the Children's Personality Questionnaire and of the Educational and Industrial Testing Service to reproduce the Extraversion Scale of the Junior Eysenck Personality Inventory is deeply appreciated.

Drs. Urie Bronfenbrenner, Earl Jennings, Lawrence Kohlberg, Milton Rokeach, and William Scott graciously acted as consultants to the project, a role which we found most helpful.

The acknowledgements would be incomplete without mention of the Joseph P. Kennedy, Jr. Foundation's support for the professorship held by the principal investigator. By establishing this position, they made it possible for the original proposal to be developed. By their continuing support, they assure that this line of investigation will likewise continue.

CHAPTER I

NATURE AND SCOPE OF THE STUDY

The Problems

Subcultural Differences in Values

It can be argued that the U. S. public schools are organized around the basic core culture typified by the middle-class suburban home. This dominant orientation is mediated by its socializing agents, the teachers. Since the schools have the communication of the culture as their major purpose, it should be no surprise that they primarily attract those individuals in basic agreement with the culture's dominant orientation. Even when the personnel are from a subculture which may not express the predominant values and attitudes of the schools, those individuals who actually go into teaching are often indistinguishable from teachers recruited from the core culture (Diener, 1967).

The consistency with which the schools represent the basic middle-class culture can be seen in the materials that they use. While these materials are becoming more diverse because of the growing recognition of the plurality of modern society, it is still true that most school reading materials portray a middle-class suburban family consisting of a non-employed mother, two or three children, a father with a white collar job, an automobile, a house with yard, and a dog. As Johnson and Geoffroy (1968) note, the significant types of values and behaviors that are portrayed in current basic readers are those generally ascribed to the white middle-class in America and allow little possibility of identification for the child who does not come from this setting. Indeed, even when minority children are occasionally represented -- and this is the exception rather than the rule -- they are seen as adopting and expressing the middle-class values (Gast, 1966). While the melting pot philosophy producing these materials has been on the decrease, much of it still lingers on.

From this perspective, children in the public school system who do not come from the middle-class white culture can be expected to experience a number of discontinuities between classroom and non-classroom settings. Such possible areas of conflict will include differences in speech and other habit patterns, differences in attitudes, and differences in ethical and moral value structures. While differences distinct from ethical values -- such as habits, attitudes, and norms of behavior -- will undoubtedly influence the child's adaptation to the school setting on a day to day basis, it is apparent that the role of value differences may also be substantial.

This position naturally assumes that children who come to the public school system from minority cultures differ in values. The extent to which this assumption can be defended is still

questioned. On the one hand, there is evidence that culturally related differences in value orientations exist which might affect achievement and adaptation in the school setting (e.g., Kluckhohn & Stodtbeck, 1961; Triandis & Triandis, 1969). Within the U. S. culture, Muncy (1967) found that high school students from the core cultural background differed in values from those from a Navaho background. In addition, Heath (1971) found differences in how black and white students interacted with white teachers, some of which might be related to value differences.

On the other hand, it appears that such cultural differences are not as strong in the area of value structures as one might think. Even school children in such divergent countries as Turkey and the United States do not appear to have distinctive value structures when the subjects are matched on such variables as father's education (Helling, 1967). And while Romero (1966) found some ten areas of difference between Spanish-American students and Anglo-American students, he found little cultural conflict in the school setting since the Spanish-American students were basically compliant with the Anglo-American values. Sundberg (Sundberg, et al, 1970) conclude from their study of students in India and the U. S. of A. that "the two groups showed four times as much communality as divergence."

The lack of a definite distinction in values between cultural groups is even more apparent when one moves to the Negro subculture within the U. S. culture. While Liebow (1967) was not objectively systematic in his research, he did conclude that his participant observation of some two dozen streetcorner Negro men questioned the usual assumption of a distinctive Negro subculture. His conclusion was that the group had the same basic goals and values that are found in the larger society but were not able to obtain them. More concrete evidence of the same conclusion is presented by Brozovich (1965). He notes that although race and sex are both shown to be related to classroom values, his hypothesis that culturally deprived children would have values in conflict with conformity and success in the classroom found little support in his data. But this general conclusion is debated by Keil (1966) who feels that a distinctive subculture has developed with particular importance for sex role identity.

Differences in values may be tied to non-ethnic background variables such as socioeconomic class differences. It has been found that value orientations do differ between upper and lower socio-economic classes (Gunderson & Nelson, 1966; Harris, 1970; Larson & Stutker, 1966; Rettig & Pasamanick, 1962). And Bronfenbrenner's (1962) review suggests that social class may be a prime correlate of differences in the related area of moral development.

<u>One of the objectives of the present study was to further examine subcultural differences in values among children in the schools</u>. Such information would be useful in evaluating the sources

of value differences and conflicts within the schools.

Note that the present project collected no data regarding the source of subgroup differences nor regarding the non-school sources of value pressures. This does not arise out of a lack of concern. On the contrary, the present frame of reference is concerned with the whole child operating within a set of overlapping environments. Peers from outside the school, parents, adult relatives and friends, sibling structure, proximity to disrupting influences, divorce, and a host of other variables are relevant to how children come to hold different values and how those values affect their lives. But a research project must draw boundaries to limit its attempts to the possible, and our boundary is the school setting.

Values and Behavior

Even if differences between subcultural groups are found, they may be irrelevant in the school setting. This would be true if values as commonly expressed, i.e., in verbal interaction or in response to questionnaires, are unrelated to behavior. Theoretically, this should not be so. Values are usually viewed as generalized attitudes of immense importance to the individual. Rokeach (1968) defines values as centralized belief structures of the individual which deal with the ultimate goals within his life. Scott considers a person to "entertain a value to the extent that he conceives a particular state of affairs as an ultimate end, an absolute good under all circumstances, and a universal "ought" towards which all people should strive" (Scott, 1965, p. 15). The fact that values consist of what a person feels as ultimately good, absolute, and a universal maxim underscores their importance for the individual and suggests that they should influence his behavior wherever the situation provides sufficient freedom.

The arguments against values relating to behavior point to many low correlations found between values and behavior. For example, Mischel (1968) builds upon the work of Hartshorne and May to point out that responses to general questions regarding such characteristics as honesty often correlate virtually zero with overt honest behavior in a particular situation. But this is not as great a problem as it seems. Theoretically, the present view of values as a generalized approach to one's life suggests that the value differences between the teacher and pupil may influence or be reflected in many situations. Whereas values may not be highly predictive of any one incident per se, they would be expected to relate to a large number of different situations. A value should therefore correlate with other variables which represent equally generalized perceptions, e.g., the total set of value conflicts over a semester, a teacher's perception of a child based on many experiences, and grades.

There is evidence that differences in values do influence a wide range of behaviors outside of the school, such as interpersonal events. For example, therapists prefer clients whose

values are similar to their own and rate these as more improved (Welkowitz, Cohn, and Ortmeyer, 1967). Value structures predict leadership behavior among school principals (Stromberg, 1967). Also, people prefer to be friendly with those who are of a different race but are similar in basic belief structure over those who are of the same race but different in values even in Louisiana (Smith, Williams and Willis, 1967). Values can also be helpful in predicting the behavior of people in new situations (Campbell, 1964). One can readily hypothesize that those students with value structures similar to their teachers will be more welcome by and more receptive to that teacher.

Evidence that verbally expressed attitudes aid in predicting school oriented behavior has also been provided. Prediction of grades over and above that which can be obtained from abilities alone has been demonstrated by Cole and Miller (1967). They found that achievement oriented values predict grades independently of abilities. Evidence that such results can be generalized from college students to students in public schools is provided by Hummel & Sprinthall (1965). They found that the values of planning and mastery of the environment differentiated between underachievers and overachievers. Of course, not all results are positive; Currie (1966) found no difference between the values of academically successful and unsuccessful children.

Values also relate to other behaviors which will often hinder or help in the educational tasks. Shotland (Shotland, et al, 1970) has found overt honesty to be related to how one rates the value of honesty relative to other values.

From the perspective of the present writers, the studies relating values to academic achievement and other behaviors of interest in the school setting are theoretically limited even though they have been statistically significant. First, they fail to recognize and take into account the complex forces operating in a particular event. Not only does a child's values affect the outcome but the child's other personality characteristics, his perceptions of the teacher's and peers' potential reactions, and his past experiences with the restraints existing in that particular situation are among the factors influencing his behavior. Second, past research has failed to consider the relationship of values to events in the school setting as an ongoing, transactional process. This latter view emphasizes the fact that both the teacher and the pupil are involved in a transaction and so both usually change as a result of those events. The transaction, however, is not an event taking place independent of other events but is part of the ongoing behavioral flow. Events are remembered and take on special meaning only when they "stick out" of the behavioral flow for the participants in that particular transaction.

The integration of these two points, the multivariate nature of causation in human events and a transactional perspective, suggests that research on values in the classroom should be

concerned with the time process, personality variables, and the other individuals involved in the classroom transactions, i.e., the teacher and peers. An objective of the present study was to relate values to reported behavior from this multivariate transactional perspective. This perspective shifts the focus of research from values in an absolute sense to values in a relative sense. The discrepancy between a pupil and a teacher is the first point of departure for analyzing how both teacher and pupil adapt to each other in their ongoing transaction. Since the adaptation in the school setting is defined by the teacher's perception of the pupil-teacher relationships over the semester, the manner in which each teacher develops her labels for the children is also of prime interest.

The Impact of Value Discrepancies

It is apparent that value discrepancies between the child and the school would have several possibilities for influencing the child's development in both academic and non-academic areas. Since friendships may be based on similarities of values, the child whose values differ more from the teacher's will probably have a more difficult time. Cox (1968) did find smaller value differences between student and teacher to be correlated with the student's perception of an effective student-teacher relationship. Since we assume the teachers hold to the middle class value system, children from other backgrounds could be expected to have more difficulties with teachers. In addition, inasmuch as the teachers are oriented towards academic achievement by the nature of their role, the child who does not value this will often be at a disadvantage. This disadvantage will be twofold: On the one hand, the child will not have the motivation to strive in the school setting. On the other hand, his values will differ from those of the teacher and so the interaction between the two will be less than optimal.

Little research has been done on achievement and value discrepancy per se, but that which has been done supports this theoretical conclusion. Bills (1952) found that college students who had the same general value orientation as their teacher did receive higher grades than others of the same I.Q. In a more detailed analysis in the public school setting, Gross (1967) examined differences in value structure between the pupils and the English teachers. The teachers, it might be noted, were so homogeneous that he was able to combine them into one profile with which to compare each pupil. He found that the pupil who had a value discrepant from one of the teachers was different from all of them. Adding value differences into the regression equation increased the prediction of the child's total English grade over that attained by abilities alone.

One manifestation of a value discrepancy between a student and his teacher may be seen in the child's overt behavior. The child who is at variance with the school system would not be highly motivated to follow the school's rules and regulations, nor would he see his

teacher as a part of his reference group. The result would be that this child could be disruptive of the class setting and be seen by the teacher as a definite behavior problem. In this manner, the value discrepancies would lead to overt value conflicts, i.e., classroom confrontations involving different value perspectives. Evidence for this position has been found from the examination of delinquents; they tend to reject middle class values and the usual means of obtaining middle class goals (Landis and Scarpitti, 1965; Scarpitti, 1965).

But not all children would be expected to act out their value conflicts. It would appear that whether or not they do so would be a function of several other variables. First, one would expect that the child's respect for the teacher as a powerful figure would be important. If the child stands in awe of the teacher's obvious ability to control the situation, little overt conflict could be expected. Second, the child's reference group of peers should influence his response. If they are at odds with the school system, then overt manifestations would be much more likely since the child would probably be socially reinforced for such behavior. This theoretical rationale would be consistant with the fact that low achievers appear to be more interested in affiliating with peers than do high achievers (Ringness, 1967).

Third, personality variables may also be influential in whether or not the child acts out a value conflict. The most likely personality variable would be that of extroversion. It has been suggested that the extrovert is characterized by a "thick-skinnedness" which allows him to develop interpersonal relations even when such relations may be viewed as painful by the more sensitive introvert (Gorsuch and Cattell, 1967). The fact that delinquents have been found to be more surgent, a characteristic of the extrovert, lends support to this hypothesis (Kelly and Veldman, 1964) as does the theory of Eysenck (1960).

If the child had a high respect for the potency of the teacher, did not have a set of peers that reinforced his disparate value structure, or was introverted, then one would expect little overt value conflict with the school system. Instead, the child would be in a situation of internal conflict. While such internal conflict is more difficult to define and measure than overt conflict, one would expect it to be manifested in the child's anxiety level, the teacher seeing the child as having personality problems, or in a shift of the child's values towards the teacher. The latter would, of course, be more likely if the peers he likes to associate with are similar to the teacher.

Regardless of whether the child turns his value differences into actual value conflict or is internally affected instead, one would expect a decrement in achievement performance. Those children with value structures similar to the school system's would be expected to achieve better when one controls for intelligence. Either an internally or externally manifested value conflict would be expected to reduce achievement.

A prime objective of the present study was to examine the extent to which overt value conflicts in the school setting are related to value discrepancies and how these discrepancies are manifested in the life of the child. In keeping with the research and theoretical considerations noted above, those children who have discrepant values will be expected to manifest them in either overt acting out (value conflicts) or in internal personality problems, depending upon their particular personality and the interpersonal situation at the time the discrepancy becomes hard to bear.

Prerequisite Questions

The Nature of Children's Values

A difficulty restricting research on children's values is the relative lack of measuring instruments. The most widely used value instrument, Allport, Vernon and Lindzey's Study of Values (1960) is inappropriate on two grounds. First, it is a theoretically oriented instrument based upon Spranger's Types of Men (1928). This theory has not, however, been extensively evaluated, and when Hunt (1968) did evaluate the religion scale of the Study of Values, he found that the characteristics which were typical of high versus low scorers were not those expected from Spranger's definition of that value. The second difficulty is that it was designed for college students and adults. While it might be used with high school students, it could not be used below that level with any guarantee of success.

Attempts have been made to build value scales which are based on distinctions made by the subjects themselves, and thus avoid one problem of the Study of Values. Scott (1965) identified numerous variations in value orientation in his investigation of a fairly homogeneous group, i.e., college students. His dimensions included intellectualism, kindness, social skills, loyalty, academic achievement, physical development, status, honesty, religiousness, self control, creativity, and independence. These appear to be manifestations of both ultimate goals and the means to those goals. His results showed that these variations in values could be reliably measured and related to various behaviors of the individual in both the academic and non-academic settings.

The procedure of seeking the values which the students themselves describe has several advantages. First, it reduces the problem of the investigators forcing their value system upon others. Each person is free to respond to an open ended inquiry; he can therefore use his own value categories and is not limited to those of the investigator. Value categories are operationalized only after they are shown to be appropriate for the groups under investigation. This would seem to provide obvious advantages for research concerned with subcultural differences in values.

The second advantage is that Scott's approach tends to

concentrate upon those values where there is conflict and difference of opinion. If the value is universally accepted by all children and adults and the child himself is well indoctrinated in that value, then it is less likely that he would talk about it. Such a value is assumed even as we generally assume the value of life. Instead, when asked for the values which are admirable in others, one is inclined to think of the current conflicts and what he would like to see changed in the near future. Responding in these terms most likely elicits value categories where discrepancies and conflicts may actually occur.

Values are here defined in the same manner as defined by Scott. This approach takes the children's values seriously and reduces the possibility of the investigators imposing their own on the situation. Scott has shown that the resulting categories were those which had the three unique characteristics by which he defined values. An additional reason for accepting his approach to definition of values comes from the nature of the values which he found. Rather than centering upon abstract generalities with which most people would agree, such as "a world at peace," his procedure turned up values which obviously have a high possibility of relating to a person's ongoing behavior and would be important for situations such as that found in classrooms.

The definition of values in the present project is not unlike that used in Ryan's research (Ryan, 1971). Both exclude opinions, beliefs, preferences, wants, and competences while being concerned with the "oughts" and "desirables." It is more distant from Guilford's (1971, p.8) definition of values as "attitudes determined by affective responses to objects, persons, activities and situations" and more similar to what Guilford labels "moral attitudes."

This approach to the definition and measurement of values in children makes the distinction between values, attitudes, life styles, and preferences quite apparent. Though it is probably true that there are many differences between the subcultures in all these areas, only values will be investigated in the present study. This is not because the other areas are conceptually unimportant but are due to the limitations inherent in any one research project.

Scales such as Scott's are not appropriate for grade school children. It appeared that little attention had been given to values as perceived by children in the public school system. Therefore, an objective of the present study was to develop an instrument for measuring verbalized values of children. Chapters 3 and 4 report our efforts to develop value scales.

Value Shifts

The values a person holds are theoretically enduring. However, "enduring" is a relative term. They should be more stable

than moods but this is far from saying that they are unchanging. One would expect an adult's values to shift if he were involved in situations for which his previous experience had not prepared him.

It is the same with a child, but there is also an additional change element: The child is still developing his values. Maturation, the impact of socializing agents attempting to shape his values, and the effect of his peers are all important and could lead to change. The research on moral development (e.g., Guilford, 1971; Peck and Havighurst, 1960; Kohlberg, 1969) does suggest that changes across time do occur.

An understanding of how children's values shift provides a background against which the impact of value differences needs to be considered. It gives an indication of whether the school does have some impact on children in this vital area. If it does, the prediction of behavior from values is necessarily limited for it does not take value changes into account. Chapter 6 reports on the analyses related to the present study's objective of analyzing when values change.

Teacher Perceptions

Since events in schools are almost always viewed from the vantage point of the teacher, this research project has also used teacher perceptions as outcome variables. But teacher perceptions are just that: perceptions. This is true even when they are turned into points along a rating scale. As such they are susceptible to halo effects and other influence such as expectations. As a result, the present project has also been interested in the teacher perception and labeling process. A further purpose is therefore to identify the extent to which teacher perceptions are a function of value differences as compared to other variables. Chapters 7 and 8 report the enquiries into this area.

The Interelatedness of the Objectives

Analysis of these preliminary questions in Chapters 3 through 7 provides necessary background for Chapter 8 which is concerned with the relationship of values to perceptions of the child by the socializing agent. But it should be noted that this is not a cumulative process of moving from Chapter 3 through 8. Rather, each chapter is examining a different perspective of the interaction between the pupils and teachers naturally involved with each other in a classroom over a semester. The unitariness of experience makes the chapter divisions arbitrary. For example, the value conflicts analyzed in Chapter 8 could also be considered as a major influence on the child's value categories such as are discussed in Chapter 3. Or again, value differences may lead to conflicts, as described in Chapter 8, which in turn change children's values, a topic discussed in Chapter 5. While an investigator must choose a perspective from which to look at the data of each chapter, it

is good to remember that the data could be fruitfully discussed from other perspectives as well.

CHAPTER II

RESEARCH STRATEGY: A FIELD STUDY

Studying Values

There are several approaches that one could take in investigating the effect of value discrepancies in the classroom. These procedures, such as experiments, surveys, and field studies, vary on several dimensions including the extent of manipulation by the investigator, the naturalness of the setting, the time element involved and the intrusiveness of the measurement. In research on values, it does not seem wise to engage in manipulation during the initial stages. It seems better ethics to begin research in this area with a non-manipulative study which leaves the situation as unaltered as possible. Such a study then lays a foundation for manipulative studies with some assurance that the manipulations will be properly successful within a broader ethical perspective.

Research in controlled settings, such as the laboratory, has much to recommend it, but such research does require a subject area where generalizability to natural phenomena can be easily assumed. Value conflicts typically occur only in a complex interpersonal setting. A laboratory would be hard pressed to reproduce the complexity of the setting, particularly when that setting is a school. Therefore, it seemed most appropriate to develop the project within schools.

Since the desire was to generalize to events in the ongoing school setting, the present research study tried not to intrude into the classroom. To a certain extent this was feasible. For example, value conflicts were gathered by having the teacher report them outside the classroom rather than by interviewing children concerning events, the teacher's responses, and so forth. But the only convenient way to gather the children's value and personality information was to give them a set of questionnaires. To the extent that these tests changed the events in the classroom, the study would have difficulty generalizing to other classrooms. It is not felt that this is a serious problem for the children; the teacher's behavior may have been somewhat more effected.

As a result of the above arguments, the project was conducted as a field study. Children and teachers were given questionnaires. Teachers also reported events from the classroom--some of which were classified as value conflicts by the staff-- and their perceptions of the children. Some of the aspects of the project required collecting data on only one occasion from a given classroom; other aspects were investigated by collecting data on several occasions across a semester. Three major field studies were conducted.

Subject Samples

Design

All the teachers and pupils involved in the study were members of intact classrooms. These classrooms were within the Public Schools of Metropolitan Nashville-Davidson County, Tennessee. Research was restricted to classrooms structured in a traditional manner, i.e., they had one teacher who worked with the children most of the day in a single room.

To obtain subjects with both an adequate reading level and less than complete adaptation to the school, the studies were conducted in the fourth, fifth, and sixth grades. The upper grades, fifth and sixth, were used whenever writing ability was needed. The lower grades, fourth and fifth, were used in the other research but only when prior checks were made to reduce the reading level requirements of the materials.

Since subcultural differences were of obvious interest to the project, samples were always formed from two categories of race crossed with two categories of socioeconomic class. To allow for a more plentiful supply of subjects and provide preliminary data on developmental questions, two different grades were used. The resulting design is given in Figure 2. Although the project design was developed independently, it is almost identical to that recommended by Cohen (1970).

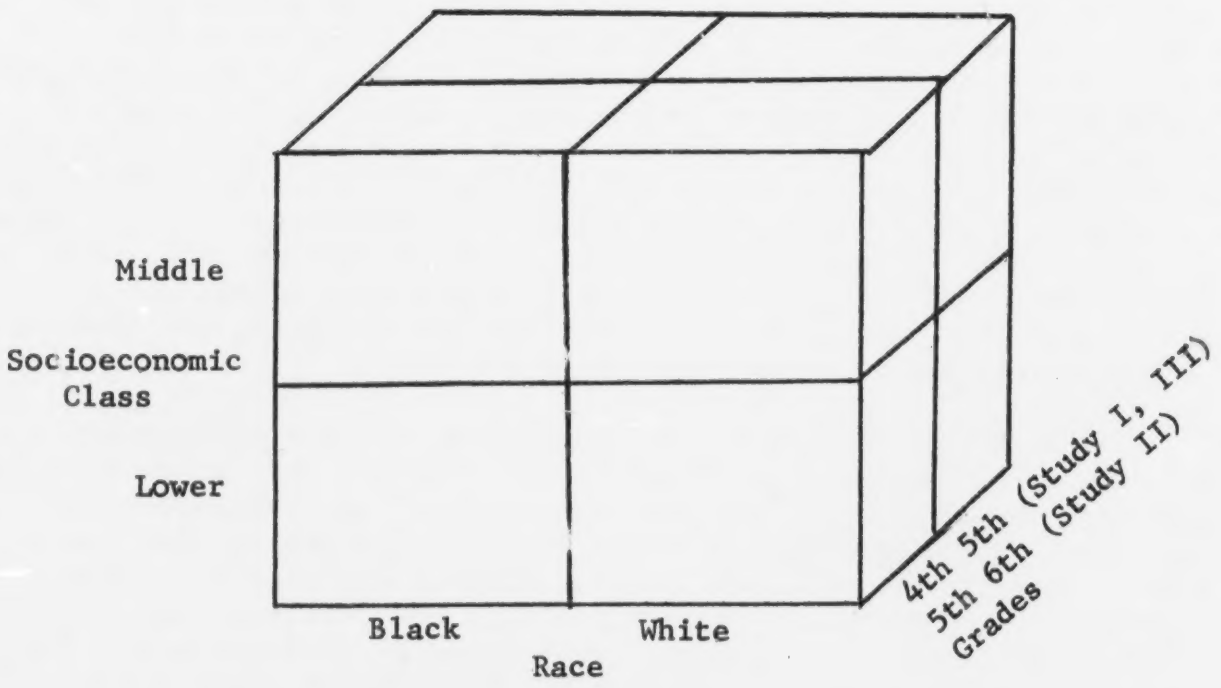


Figure 2.1. Sampling Design.

Studies I, II and III

Three major studies were conducted under this project. Study I's data consisted of written responses to open-ended questionnaires and formed the basis for Chapter 3. Study II was concerned with data for scale evaluation. The analyses of these data are reported in Chapter 4. Study III provided the data for the analyses of prime interest. In this case children and their teachers participated in data collection procedures across the fall semester.

Tables 2.1 and 2.2 summarize the characteristics of the sample used in each study. The first sample used only a few racially integrated schools. The next two samples used only schools serving racially segregated neighborhoods. In the third sample, the schools were selected so that there were two schools to represent each of the race and class cells in the design. One fourth and one fifth grade classroom was involved in each school in that study.

Table 2.1

General Information on Samples Used

	Study I	Study II	Study III
Time of Testing	October, 1969	May, 1970	September, 1970 to January, 1971
No. of Schools Tested	3	5	8
No. of Classrooms Tested	13	24	16
Racially Integrated	Yes	No	No
Grades Included	5th, 6th	4th, 5th	4th, 5th
Total No. of Usable Subjects	256	522	397

The number of classrooms in each study was determined by a rough estimation of the number of subjects needed for the analyses planned. The second study involved the largest number of subjects since it evaluated an extensive item pool. The possibilities of capitalizing upon chance are great when individual items are being examined and the N needed to be larger than was necessary in the other studies.

Table 2.2

Sex, Race and SES Composition of Samples

		Study I		Study II*		Study III*	
		Race		Race		Race	
		B	W	B	W	B	W
Social Class	Lower Class	Males	54	35	85	69	39	12
		Females	56	38	114	58	56	24
	Middle Class	Males	19	16	47	52	38	71
		Females	21	17	50	47	48	89
Totals			256		522		377**	

*SES based on teacher's estimate.
**20 subjects whose SES was unknown are not included in this table.

Bias Analysis

In Studies I and II, all pupils in each classroom participated in the study unless they were absent on the day of testing. While each child was told he did not have to answer any of the questions which he did not want to, the children answered all of the items in almost every case. Therefore the bias in the samples was deemed negligible and assumed to be randomly related to the parameters of interest in those studies.

In the third study there was some possibility of bias being introduced by selective participation factors. Bias in sampling arose from four possible sources. First, though two later attempts were made to test anyone who was absent on the day when the rest of the class was tested, this was not sufficient to pick up all subjects. Subjects were dropped from the analyses because they were absent during either the first or the second testing session. Second, several children were also "absent" from the second session because they had moved out of the school tested. Third, some children did not participate in the study because their parents refused permission for them to do so. The fourth group not included in the study were subjects who did not return permission slips.

For the pupils in each of the four categories above, some information was available. This consisted of data reported by the teacher and included the pupil's sex, race, socioeconomic status as

rated by the teacher, teacher's estimate of the child as a "good student" on the Self Anchoring Scale, and the number of both positive and negative critical incidents in which the child was involved. At this point, we had somewhat of an ethical problem. On the one hand, the data which had already been collected without participation by the children would be useful in analyzing the characteristics of the sample; on the other hand, consent for the subjects' participation had not been given.

We decided to use the data on three grounds: First, these subjects had not provided these data and therefore they were not imposed on. Second, the data were deemed to be the property of the school system as much as the property of the children and the school system had given permission for its use. Third, it was deemed of considerable importance to understanding the generalizability of this study. Since the third point seemed to outweigh any possible violation of the students' rights, the data analysis comparing nonparticipants and participants was undertaken. Data on nonparticipants were not included in any other analysis.

A multiple discriminate analysis (Nunnally, 1967) was used to see whether the four groups of pupils not included in the study and those forming the subjects for the study could be distinguished on the basis of their scores. The prediction variables included sex, race, socioeconomic status (SES) of school, SES as rated by the teacher, the number of positive incidents in which the child was involved, the number of negative incidents involving the child, and the teacher's rating of each child on the Self Anchoring Scale. These measures are described in more detail later.

The first two discriminant functions were found to be statistically significant ($p < .05$) and distinguished the groups. Figure 2.2 plots each group centroid in the two dimensional space. It can be seen that the first dimension separated out the subjects in the study not only from those who moved during the semester but especially from those who did not return the permission slips. The first dimension correlated .9 with the Self Anchoring Scale and is essentially that scale. The first dimension also correlated .4 with race and sex; this indicates that white females tend to be more highly represented in our study as compared to the group for whom permission slips were not available. An examination of this figure therefore shows that some bias was introduced and that those subjects whom the teacher felt were her worse pupils tended to be eliminated. Fortunately, this dimension did not correlate with involvement in classroom incidents.

The second dimension in Figure 2.2 separates out those whose parents refused to permit them to participate in the study from the other groups. This dimension correlated .7 with race, which shows that the group whose parents objected to their participation in the study were primarily white.

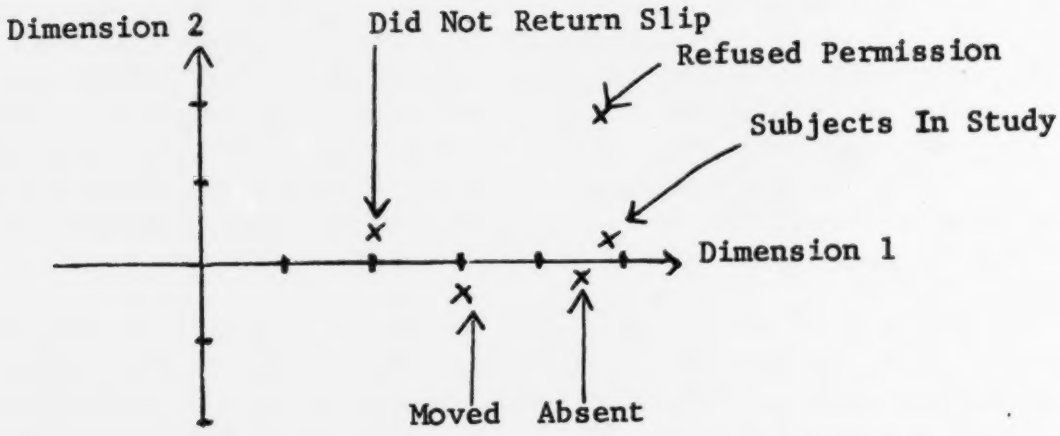


Figure 2.2. Separation of the subjects and non-participants in two dimensional space.

The second step of the bias analysis was to follow these overall significant tests with analyses of variance and t-tests as appropriate. It was of particular interest to compare the total set of children not in Study III with the total set of subjects in the study to evaluate the overall bias. The results of the t-tests on these seven variables are given in Table 2.3. The last two columns show that, as expected from the first dimension's multiple discriminate function, the participants in the study included relatively more females than did the combined groups of non-participants. Since the two multiple discriminate functions operated in opposite directions with regard to race, the total sample is not biased on that variable considered alone. Significant differences with the participants having more positive critical incidents than the non-participants and also being rated higher on the teacher's Self Anchoring Scale are functions of the first discriminant function.

Note that on neither measure of socioeconomic status was there an overall difference between the participants and the non-participants. In individual t-tests, the no return group were rated by the teacher as having slightly more lower class children than the other groups. The no return rate was, however, the same for all school SES groups.

Since there was a significant bias effect in the sample, the results of this study should not be generalized too far. It can not be applied to those children whose parents would raise serious objection to the study nor to children who would either be negligent in returning the permission slips or whose parents would not fill them out.

Table 2.3

Means of the Participants and
Non-Participants on Available Data

	All Pupils N=502**	Non-Participants					Subjects In Study N=397
		No Return N=33	No Permission N=28	Moved N=25	Absent N=19	Total N=105	
Sex (1= male 2= female)	1.53	1.30*	1.40	1.46	1.47	1.39*	1.57
Race (1= black 2= white)	1.50	1.27*	1.92*	1.43	1.53	1.50	1.50
Estimate of School SES (1 = middle 2 = lower)	1.50	1.50	1.32	1.57	1.63	1.50	1.50
Teacher Estimate of SES (1=middle 2=lower)	1.29	1.50*	1.08	1.39	1.36	1.35	1.28
Positive Critical Incidents	.50	.28*	.32	.46	.38	.35*	.54
Negative Critical Incidents	.77	1.03	1.76	.78	.15	.89	.74
Self Anchoring Scale	6.19	4.47*	6.32	4.75*	6.42	5.37*	6.41

*$p < .05$ between this group and the subjects in Study III.

**No significance test run.

Procedures

Data Collection

For each of the studies, the schools were selected in consultation with the Metro school representatives. The previous year's

report of the distribution of students by race and grade within each school was examined to find those schools which could supply the necessary numbers of subjects. A further examination of the schools was then made to determine whether or not they could be placed in a lower or middle socioeconomic category. The first study utilized the school's reputation to place it into a socioeconomic category. The second study also involved selecting schools by a similar method.

Since the third study involved the heart of the entire project, greater pains were taken to find schools appropriate to the design. A school was placed in an SES category if and only if a Board of Education social worker, the Metro Schools Research Division, and the principal himself placed it in the category. Asking the principal what SES most of his students were was a variant on the procedure Kohn had found advisable in his national study of principals of schools and research on social class, values, and child rearing (Kohn, 1969).

Testing in the classrooms was carried out by the project's staff. The teacher introduced the staff person and then left the room. The research person explained briefly the purpose of the project and assured the children that no information on an individual student would be given back to the schools. The research instruments were then read aloud in each classroom in all the studies. If time permitted after the testing, a brief open-ended discussion would ensue with the children during which their reactions to the materials would be discussed in a permissive, accepting atmosphere. In all except a few cases, a second person was in the classroom with the person administrating the materials to provide aid where necessary.

In Study III, the testing materials were given in two sessions and the halves were counterbalanced by order of administration. That is, the testing materials were divided into Parts I and II. Half of the subjects took Part I first and the other half took Part II first. Counterbalancing was by classroom, with classrooms so assigned that half of the pupils in each cell of the design took the materials in each order.

In Study III the first testing session was fairly long for this young a child. It was noted that some children appeared to become tired and to do poorly on the non-verbal test of intelligence when it was given last. Therefore, this and several other tests were checked for order effects. No effects that would influence the results of this study were found.

No testing session lasted more than 1 1/2 hours without a break. All materials prepared by the project were written at a fourth grade or less reading level.

In researching values on children, it could be that the

race of the person administering the instruments would have some effect on the outcome. In the first two studies, the race of experimenter was completely counterbalanced. Analyses of the first study's data and of an independent project (Savarese, 1970) led to the conclusion that the race of the administrator had no effect on the measures used in this study (cf. Appendix A).

All materials used in each of the three studies were cleared both through the Office of Education and through the Peabody Committee on Research Ethics. In addition, the school officials reviewed the materials and approved them. While no principal or teacher was required to participate in the study, most of those contacted did so. Some black principals in low SES schools refused, claiming that they are constantly studied, but nothing is done to improve the school conditions. In the first two studies the school system deemed the materials relatively innocuous and the study sufficiently important so that they assumed full responsibility for granting permission for the children to participate in the study. In the third study, only those children who returned a permission form (Appendix B) signed by their parents participated in the full study.

While names were gathered on the questionnaires so that children's answers could be matched with teachers' ratings of each child, identification codes were substituted for names at the earliest possible time. Materials where the child could be uniquely identified were not open to any person outside the research staff. All materials were only used for research purposes.

Measuring SES

Measuring social class by the school's reputation primarily assured some variation of the SES of the students in the study. Naturally, it was the exceptional school where all the students were from only one class. For a measure of each child's social class, two techniques were tried in Study I. The first approach was to provide a local credit bureau with the names of the parents of the pupils. They then placed each of the fathers into one of Hollingshead and Redlich's (1958) seven occupational categories. Socioeconomic (SES) ratings based solely on occupation have been shown to correlate strongly with the actual social class (Hollingshead & Redlich, 1958). Occupational categories 1, 2, 3, and 4 were grouped together as middle class since these represent the general description of "white collar" occupations (e.g., professionals, owners and managers of businesses, clerks and technicians). Categories 5, 6, and 7 were grouped together as characterizing lower class or "blue collar" workers (e.g., skilled, semi-skilled, and unskilled manual employees). The use of credit bureau ratings had two drawbacks: they were expensive and required the parents' names from the school records. Parents names were not available for ten percent of the sample in the first study.

The second approach to measuring children's SES was to use

teacher's ratings. Each teacher placed each pupil into one of three categories: "middle class," "lower class", or "don't know". These categories were not further defined. They were deemed sufficient since the sample was known to contain few upper SES students and since the raters were not given the training necessary for finer discriminations. This procedure was found to be as accurate as any known procedure which does not rely on direct information concerning the parents (cf. Appendix C). This approach was also used in Studies II and III.

When the teachers in Study III rated their students as to their socioeconomic class, the SES of the children was somewhat different than when that SES was determined by school board opinion. These tabulations are presented in Table 2.4. Note that the middle-class children outnumbered the lower class children in all of the schools. Those schools deemed lower class by the school board had a somewhat greater number of lower class children and so the relative differences were correct. It should be also noted that schools judged as lower class were the ones with more obviously inadequate physical facilities.

Table 2.4

School Reputation and Teachers' Estimation of SES

		Teacher's SES Estimate			
		Middle	Lower	Don't know	
School SES	Middle	72.8% (142)	22.3% (44)	4.9% (11)	100% (197)
	Lower	52.0% (104)	43.5% (87)	4.5% (9)	100% (200)
				TOTAL:	(397)

Since teachers' estimates are probably more accurate than the rater's judgment of the school, this measure of social class was generally used in the third study. When teachers were not able to estimate the SES of a child, they were given their school's rating.

But on certain occasions it was not possible to use teacher's estimates of SES. When equal numbers of subjects were needed in cells for analysis of variance purposes, the lower percent of lower class pupils would often lead to too few subjects for the analysis

to be run. In this case, we used school rating of socioeconomic status. While the school rating was not as accurate and therefore could not lead to significant results as often as the other measure, significant results still reflected class differences. Insignificant results could, of course, arise solely from the inaccuracy of measuring social class of the students by the schools reputation but this was felt to be better than running no analysis at all. The other occasion for using SES of the school was when the dependent variable was a teacher rating. Since the teacher's rating of SES may be altered by her other judgments of the child, it would have been contaminated for those analyses.

Testing Materials

The testing materials for the first two studies are straightforward and described along with the analyses of the results.

Study III's testing materials are listed in Table 2.5 along with when they were given. The testing materials were of several kinds. The value measures developed in this study formed one group of scales, measures of the children's personalities another, and questions about the children's peer groups a third. The fourth category of variables consisted of those where the teacher provided information on each child.

The Value Socialization Scale, the Semantic Differential Value Scale, and the Rank Ordering Value Scale are paper and pencil instruments. They represent three different ways of measuring the basic value dimensions as identified and operationalized in Chapters 3 and 4. Since these are questionnaires, there was a possibility that the responses may have been distorted by children attempting to put themselves in a good light. To allow for checks to be run on this response hypothesis, the Children's Social Desirability Scale was also included.

The scale to measure the perceived potency of the teacher was semantic differential in form. The concept rated by the children was "my teacher." The bipolar semantic differential adjectives scales found to be most productive in building this scale were strong-weak and dull-sharp.

The peer group was defined as those children within the classroom with whom they "ran around" during the semester. This was operationalized by asking the child to list those children with whom they went around, those with whom they would most like to go around, and, at the end of the semester, those with whom they went around during the semester. The latter was the primary measure used in the analyses.

Table 2.5

Basic Measurements: Study III

	Testing Occasions: Sept.	During Semester	January
I. Teachers' and Children's Values (cf. Chapter 4 for detailed description)			
*Value Socialization Scale	X		X
*Semantic Differential Values Scale	X		X
*Rank Ordering of 8 Values	X		X
II. Child's Personality			
Extraversion-Introversion (Eysenck, 1965)	X		
Culture Fair Intelligence test (Cattell, 1950)	X		
Verbal Analogies (Scale B of CPQ, Porter, Cattell, 1963)	X		
Anxiety (Preliminary children's version of the STAI; Speilberger, Gorsuch & Lushene, 1970)	X		X
Children's Social Desirability (Crandall, Crandall, & Katovsky, 1965)	X		
Children's Locus of Control (Bialer, 1961)	X		X
*Perceived Potency of Teacher	X		X
III. *Child's Peer Group			
Who they go around with	X		X
Who they would most like to go around with	X		X
Who they went around with during the semester			X
IV. Teacher's Ratings of Child			
*Quality as a student: Self-Anchoring Scale (cf. Chapter 7)	X	X	X
*Values: Academic Achievement, Affiliation, Good Manners, Physical Appearance and Development, Obedience, Kindness, Honesty, Diligence	X	X	X
Personality Problem? (Peterson, 1961)	X	X	X
Behavior Problem? (Peterson, 1961)	X	X	X
Grades (Average of those grades which the teacher gave)			X
*Socioeconomic Status (SES)			X
*Important Incidents (cf. Chapter 7)		X	

*A copy of each of these instruments is included in Appendices D and E.

A major teacher rating of the child was the Self-Anchoring Scale. The teacher was first asked to give a short verbal description of what she considered to be the best and the worst students to form the rating anchors. She then rated each child on a 9 point continuum formed by her two definitions and these ratings were the scores on the Self-Anchoring Scale. The teacher also rated each child on the basic values identified in Chapter 3 as important for this study. This was a simple 9 point rating. The definitions of each value were an outgrowth of Chapter 3 and consisted of actual statements used by children from which the value categories were developed.

During the semester, each teacher was asked to report any major incidents which occurred in her classroom. These critical or important incidents were simply defined as those which stood out in her memory as she looked back over the past day or several days. The teacher was asked to indicate who had been involved, what had happened, and how she had handled it. These were picked up from each teacher on a once a week basis.

Several derivative measures were used extensively in this study. Value differences or value discrepancies were empirically defined by subtracting the child's value score from his teacher's value score and from his average peer group's value score. Since such a measure is a linear composite of two other scales in the study, any significant effects found with it could be a function of either the child's values or the teachers/peers values. Checks were necessary in order to clarify the source of relationships with the discrepancy measures. Although from a technical viewpoint the inclusion of these measures was not necessary, their heuristic value in communicating both the perspective of the study and the results was an overriding consideration. (Note that these are not difference scores used to measure change.)

Table 2.6 gives the internal consistency reliability of each scale for each occasion when it was given. Most of the reliabilities are adequate for research purposes with large numbers of subjects. Only the Extraversion scale is questionable and the Locus of Control scale inadequate (a second administration of the latter with more carefully developed testing materials did not improve the picture).

Several of the instruments used in the third study were examined to determine whether or not their reliabilities varied as a function of verbal ability. This was definitely the case with the Locus of Control scale (Bailer, 1961). Those subjects with poor reading levels measured, by their verbal ability (Scale B of the CPQ, Porter and Cattell, 1963), had unreliable scores. Therefore this measure was dropped from the study. The other instruments were deemed satisfactory by this criterion (cf. Appendix F).

Table 2.6

Reliabilities: Study III

	Internal Consistency Reliabilities			
	Childrens'		Teachers'	
	Fall	Winter	Fall	Winter
I. Teachers' and Children's Values				
Value Socialization Scale	.77	.76	.68	.61
Semantic Differential Values Scale	.95	.95	.96	.97
Rank Ordering of 8 Values	-	-	-	-
II. Child's Personality				
Extraversion-Introversion	.47	-	-	-
Culture Fair Intelligence Test	.74	-	-	-
Verbal Analogies	.67	-	-	-
Anxiety	.72	.84	-	-
Children's Social Desirability	.90	-	-	-
Children's Locus of Control	.24	.42	-	-
Perceived Potency of Teacher	.64	.66	-	-

Missing Data Procedures

In Studies I and II, response to incomplete items were supplied by a simple random choice among appropriate options. In Study III, however, this problem was handled by randomly choosing the response made by a same sex subject in the same classroom to the item at issue. In those cases where entire scales were missing, the mean for the same sex classmates was used. In Study III, three subjects were dropped because they lacked too many scales.

Statistical Analyses

The prime multivariate procedure used to analyze the data was analysis of variance. Depending upon the dependent variable being analyzed and the purpose of the analysis, the ANOVA's had as many as six factors. The number of subjects in each ANOVA varied even within the same study. For each analysis, the largest possible number of subjects per cell was used. Proportional cell analyses were also used where appropriate computer programs were available.

The ANOVA approach, however, had some limitations. First,

variables which are non-linear in nature would need to have been arbitrarily divided into categories to proceed. Second, analyses involving variables which are highly correlated could not be run since it was practically impossible to develop analyses with equal number of subjects, or even proportional numbers of subjects, in each cell. But due to the widespread understanding of analysis of variance procedures and the fact that they do test the linear model adequately when the data can be divided into categories and are not highly correlated, this was a primary method used to analyze the results.

Multiple regression procedures were used in addition to ANOVA's to test linear models specifically designed for the hypotheses under consideration (Bottenberg & Ward, 1963; Kelly, Beggs, & McNeil, 1969). F-tests were generally computed between multiple regression equations which contained the variable or variable set being tested and the same equation without that variable. If a significant difference was found, then that independent variable added to the prediction of that dependent variable. By following a logical set of such analyses, every test normally run in an analysis of variance could be run by a multiple regression procedure (Cohen, 1968).

However, multiple regression analysis is seldom used to test all the hypotheses tested by analysis of variance. Rather, a shotgun approach is avoided by focusing on the hypotheses of direct interest to the study. This means there is less capitalization upon chance. In a fiveway analysis of variance which runs, for example, 31 independent significance tests, at least one could be expected to be significant by chance at the .05 level. But generally the investigator will not have hypothese for all those 31 significance tests nor would they be necessarily meaningful. With multiple regression analysis, the investigator simply runs those which are important for his case. Then the ordinary probability level is an appropriate description of the statistical significance of his results.

It should be noted that the advantages of the multiple regression approach to analyzing the data are several in addition to a more exact test of the hypotheses than is otherwise available. A regression analysis is comparable to a fixed effects analysis of variance which does not require equal numbers in each cell. This allows one to investigate the groups actually occurring in public schools and not to be concerned with all possible groups which might occur in some theoretical culture. It utilizes more information from the continuous variables since it does not require splitting the subjects into artificial groups. Since it does not investigate all possible interactions but only those which are previously hypothesized, it utilizes relatively few degrees of freedom in comparison to the number of moderators, covariates and variables; this means that with a reasonable number of subjects

one can expect a reasonable number of degress of freedom for any particular test of significance which might be of interest. The analysis would have the further advantage of giving some indication of the size of the effect of the variables. The results from the analysis can then be presented graphically by computing means for various sub-groups.

The major disadvantage is that compared to analysis of variance techniques, the tests of significance that are run will not be completely independent. This is a function of the inequality of sample sizes across groups. This is not felt to be a major disadvantage for an exploratory study such as this. Since the multiple regression approach requires one to test only explicit hypotheses rather than to run automatically all possible significance tests on all possible interactions, the probability of a spurious conclusion may be less than in a conventional analysis of variance approach (Cohen, 1968).

To keep the project report reasonably brief and understandable by its potential audience, individual analyses are not described in detail. Neither are summary tables of the analyses of variance or regression analyses generally given within the body of the report. Instead, the report presents those graphs and tables which show the results found along with the significance levels of those findings. All results discussed were significant by the usual statistical criteria.

Design Limitations

Where the hypotheses of the relationship of value dissimilarities to the various conflict manifestations have been upheld, one is able to conclude that a correlational relationship exists between these variables. The present design eliminates the hypothesis that the conflict manifestations examined here lead to the difference in values found since the measurement of value similarity occurs prior to that of the conflict manifestation. The present research design also eliminates most hypotheses based upon maturation, differential testing, or regression to the mean effects. (Campbell & Stanley, 1963).

However, the research design does not eliminate the hypothesis that another variable causes both the value dissimilarity and the dependent variable if that other variable is not among those included in the analysis. One would expect that individuals differing in values would have a long differential history and any part of this difference in history could be a critical causative factor in the producing of a score on the dependent measure. Another limitation of this research design is that the effect of the testing on all the students and teachers is uncontrolled. It is felt that this factor is probably minor since the testing was done as an "outside research project" and involved only a few hours of the students' time.

While this design does not allow for a causative experimental analysis, it does provide some necessary prerequisite data for such analysis. Except where this study found significant results, it would hardly be worthwhile to establish experimental procedures to test the hypothesis of the effect of value dissimilarity upon conflict manifestations. Where this study did find significant results, then this would be a basis for the possible experiments on questions of basic interest. Since this latter study would involve long range, programmatic changes in a school setting with similar, equivalent but non-value related changes in another equivalent school setting, significant results in a study like the present one should be a prerequisite to such an experimental analysis.

CHAPTER III

CHILDREN'S PHENOMENOLOGICAL VALUE CATEGORIES

On Developing A Set of Value Categories

For this project, values were defined in the same manner that Scott defined them: "A person may be said to entertain a value to the extent that he conceives a particular state of affairs as an ultimate end, an absolute good under all circumstances, and a universal "ought" toward which all people should strive" (Scott, 1965, p. 15). But this definition does not, of course, say what the values are in terms of content.

There are several possible approaches to defining the set of values in which one is interested. First, he could simply assume (or argue for) one given set as critical. Spranger (1928) produced one particular set which Allport, Vernon and Lindzey (1960) attempted to operationalize. Second, one could survey the literature and use a set of judges to identify and match the different value categories used in that literature. This is the approach of Ryan's Hawaii project (Ryan, 1971). Third, it is possible to go to the children themselves. This phenomenological approach was, for example, taken by Barthol (Barthol & Bridge, 1968; Barthol & DeMille, 1969) when they asked subjects for "good-things and bad-things." As Guilford (Guilford, et. al., 1971) indicates, this approach has a long history.

From the present perspective, the first two approaches are somewhat limited. The first one, using a priori definitions of values, means that research is conducted on only those values which the investigator feels are important. It may completely miss the value categories which are of real concern to the children or teacher, or which come from a tradition which is outside the investigators' life space. The second procedure is better; it helps the investigator transcend his own limited experiences by introducing the experiences of others. This also reduces the possibility of his own values biasing the results. And where those "others" include people from the same background as the subcultures of children in the research population, as is the case with Tryan's research, it helps to reduce the possibility of missing a category because it is peculiar to one particular subculture. But note that the views solicited are generally those of adults. This is fine if the emphasis is on differential expectations of the school as a socializing agent, but is more limited when the emphasis is on the ongoing classroom process. When the concern is with a particular setting, the "experts" to be consulted are obviously those in the setting itself. By going to school children one can develop a set of value categories which reflects their values and which reflects the values which the socializing agents are currently communicating to them. If value discrepancies occur which influence the child's classroom behavior, they are most likely

to be in the areas which the children see as value ladened. The only exception to this would be when the discrepancy is over a value which the socializing agent is attempting to communicate but which the children neither see nor understand. This problem is not likely to be major if slightly older children are questioned to identify the value categories than are used in the study proper. At least some of the older children would be understanding and using a value category even if it was difficult for them a year earlier.

Scott (1965) has used a procedure similar to that just outlined with college students. He asked them for the kinds of behaviors that others admired in a person and was able to show that such behaviors tend to be perceived as ultimate ends, absolutely good, and as universal "oughts". Scott's approach has an advantage over that of Barthol's for the present study since it narrows the area. No longer need the project include all of a person's interests, attitudes, and preferences; rather it became solely concerned with those which were socially and ethically oriented. These would be theoretically expected to be most crucial for interpersonal relationships.

The present chapter is a report of our attempt to identify the naturally occuring value categories that children themselves use. The research used Scott as a basis from which to proceed and was concerned with identifying phenomenological value categories as used by the several subpopulations of the current study. The next chapter reports our attempts to then measure each child's commitment to a particular value.

Pilot Study

Development of the Free Response Form

To illicit value oriented statements from children, it was necessary to have a series of questions to ask them. For convenience, it was decided the main study of value categories would utilize written questions. The pilot study was designed to examine potential questions with children as well as to develop a set of rating categories by which the responses to the questions could be grouped. The questionnaires were seen as eliciting free responses to broad ranging questions so the categories of the investigators would have as little impact as possible.

The questions for the initial questionnaire were developed out of those used by Scott (1965) with college students. One of his typical questions was: "Think about the various people you admire, and try to reflect on what it is about them that is admirable. Now consider the general question: What is it about any person that makes him good? What personal traits would you say are particularly admirable?" (Scott, 1965, p. 17). Scott's questions

required abstractions from the situation. Not only is such abstraction more difficult for children, it would also be somewhat self defeating for the present study. The interest here was in those value categories most appropriate to children in their experiences with socializing agents rather than the more general categories sought by Scott. Therefore the questions were revised to be concerned more specifically with the norms a child would see as being espoused by those around him.

Pilot data on a set of potential questions was collected by a series of individual interviews. Staff members went during the summer to local community centers in lower class black and white communities and talked to the available children. A total of 39 children participated, with 23 being males, 16 females, 20 black, and 19 white. These relatively unstructured conversations took from 15 minutes to 45 minutes for an individual child.

An examination of responses children made to the questions suggested that some were more efficient than others in producing value related responses. These included:

> What do children do that makes their teachers' happy? What else?
>
> Think of a girl your friends dislike. What do they dislike about her? What else?
>
> Think of a child your friends like. What do they like about him or her? What else?
>
> What kinds of things do you do that make you feel proud of yourself? What else?
>
> Think of a boy your friends dislike? What do they dislike about him? What else?
>
> What do children do that make their mothers and fathers happy? What else?

Note that these questions are of a relatively impersonal nature. It is not assumed that they reflect the values of the child but only the value categories which he communicates in. This is in keeping with the basic purpose of this value category analysis: We wanted to find those values which children understand so that we could communicate with them. We left it to Studies II and III to identify when the values were found in individual pupils or in groups of pupils.

Preliminary Value Categories

The second phase of the pilot study developed preliminary

value categories that could be used with some assurance of relevancy to children's responses and of reliability. The training procedure for raters were quite simple. They were given a copy of Scott and told to examine the value categories he found among college students. Particular reference was given to the scales which Scott developed and to the raters being able to place the items in their appropriate value categories.

The raters were instructed to examine each of every child's comments. If its meaning fell within one of Scott's categories, then the comment was reported as within that value area. If it did not fit any of Scott's value categories, then they were to use other categories as they deemed fit and to give them their own labels. This procedure gave a basic starting point which encouraged the categories being compatible with those found in older adolescents. But by giving the raters the opportunity to introduce other categories, it also allowed for new value categories to appear which might be peculiar to a particular ethnic group or to children.

Raters were college students and generally from middle class backgrounds (as were most of the staff involved in the total project). It may well be that the value labels such raters selected would not be the labels children would use. This does not seem important since the result was those value categories which grouped together the children's responses and have high content validity. Indeed, it could be argued that the value categories that researchers find meaningful are more important for the scientific community than pure lay categories would be. It does seem that the raters had little problem in placing most responses in a value category with a label close to that which might be used by a child.

Table 3.1 gives the initial value categories which the raters felt would be useful over and above Scott's 12 categories. All categories are included which were suggested at least once by one of the three raters. From this set of initial value categories, discussions were conducted until the raters agreed upon a more limited set of value categories.

After discussing the various categories proposed, the raters agreed upon a common set of categories. These categories are listed in Table 3.2. Each rater then examined all of the responses and indicated which value category was most appropriate for describing each one. The number of times a particular child used a particular category was his score in that category. The score given each child by rater 1 was then correlated with the scores given by raters 2 and 3 and raters 2 and 3 were likewise correlated together. These reliability coefficients are given in Table 3.2 which also gives a total reliability estimate.

Table 3.1

List of Initial Value Categories

Used in Addition to Scott's

1. Achievement
2. Affection*
3. Affiliation
4. Assigned tasks*
5. Competence
6. Conceit (modesty)*
7. Cooperation*
8. Courage*
9. Destruction*
10. Discipline
11. Experience*
12. Fate*
13. Happiness*
14. Imitation
15. Materialism
16. Morality
17. Neatness
18. Pride*
19. Punishment*
20. Respect*
21. Responsibility
22. Sex role
23. Trouble
24. Vocational achievement

*Used only one or two times

Based upon these re-rating experiences and the reliabilities, the value categories were discussed further by the raters and re-defined. This final redefinition is presented in Table 3.3. These are the categories which the raters felt would be adequate for covering practically all of the children's responses. The table also includes the 12 value categories of Scott and indicates where they were combined and/or subdivided in order to fit the present data more adequately.

In Table 3.3, each value category is defined by a set of examples. While occasionally the set of examples begins with a statement by the staff to clarify the category, all other responses are verbatim from the children's interviews. Note that both positive and negative incidents regarding the same value are included in the definition. For example, "To be good at sports" and "To be puny" were both placed in the Physical Appearance and Development category since they are functionally concerned with the opposites of the

Table 3.2

Rater's Reliabilities

Value Category[a]	Percent of Usage	Reliabilities			Total Reliability[b]
		Raters 1 & 2	Raters 1 & 3	Raters 2 & 3	
Intelligence	4	.43	.50	.55	.75
Kindness	23	.49	.45	.61	.74
Social Skills	14	.33	.29	.29	.55
Academic Achievement	10	.29	.55	.23	.55
Physical Appearance and Development	2	.43	.42	.36	.68
Status	3	.15	.33	.11	.35
Honesty	2	.54	.42	.42	.68
Religiousness	1	.62	.58	.54	.81
Self-Control	5	.25	.08	.08	.21
Aggression	4	.48	.32	.33	.60
Obedience	11	.72	.56	.57	.80
Diligence	7	.35	.41	.42	.68
Unclassifiable	11	.19	.12	.10	.32

[a]Categories used less than one percent of the time are not included. Responses were divided into statements before being given to raters.

[b]Calculated by correcting the median correlation between raters by the Spearman-Brown formula.

Table 3.3

Children's Value Categories

Scott's 12 Value Categories	Our Final 15 Value Categories	Exemplar Definitions
Kindness Social Skills	Kindness (passive)	Being the recipient of kindness; he is nice to me; brings me candy; smiles at me; helps me.
	Kindness (active)	Help her out; help them because I want to; he is generous; help people when they are sick; considerate person; sharing; clean the car.
	Good Manners	To meet others' social expectations; to not "cuss;" to say thank you; not lose your temper; act smart way she acts.
	Affiliation	To like to be with a person just for the sake of being with him; to like being in a group; to like talking with others.
Intellectualism Academic Achievement	Academic Achievement	Getting good grades; working hard in school; education necessary for success.
Physical Development	Physical Appearance and Development	To be big and strong; to be good in sports; to be good looking; to be puny; to not like exercise; be sloppy; shooting rabbits.
Status	Status	To be rewarded for accomplishments; being respected by others; being in a position to tell others what to do; to have a good job and a good home.
Honesty	Honesty	Never cheating; always telling the truth; informing authorities of all wrong doings; stealing; telling lies to protect a friend.
Religiousness	Religiousity	Always doing what the Bible says; attending church service; saying prayers regularly; telling others about the Bible; not believing in God and/or the Bible.

Table 3.3 (Continued)

Scott's 12 Value Categories	Our Final 15 Value Categories	Exemplar Definitions
Self-Control	Obedience	Do what they tell me; obey them; listen to them; mind them; remember them.
	Mean	Non-physical agression (as opposed to harm); she is mean; she teases; to fuss at someone; puts kids in corner; takes them to principal.
	Harm	Throw rocks at me; fight; hit.
	Diligence	Be a good worker; do their work right; save up all the money; study hard; whatever you are doing, do it the best you can.
Nonparallel Categories:		
Creativity	Good	"To be good;" be a responsible person; don't do anything bad; do dirty stuff.
Independence	Imitation	Be like my mother.

same value pole. The former response, of course, generally occurred in answer to a question about what makes a person admirable whereas the latter type of response generally occurred to a question concerned with what makes a child disliked.

Study I: Collecting and Categorizing Value Statements

To the revised set of questions generated by the pilot interviews, two questions were added:

> Why do children go to school?
>
> Think of one most important thing that grown ups should be or should do. What is it?

The former was used as a warm up question to encourage the children to write freely whereas the latter was a final exploratory question which was not scored for these analyses. These materials were given to the subjects of Study I who were described in the second chapter of this report. The questions were printed on paper with space for the children to respond, and read aloud to the children in their classrooms but with the teacher absent.

After identifying information such as name and sex had been replaced by an arbitrary code, raters were given these sheets to

score. The sheets from the various classrooms were randomly mixed for each rater so that a lower class white male might, for example, follow a middle class black female. The raters had no knowledge of the child's SES, sex or race. The raters were instructed to examine each separate response and to place it in the value category which seemed most appropriate given the context of the question and answers. The exemplar definitions given in Table 3.3 were used to define those categories. Each scorer was also told to add a category for any statement which indicated a value not already mentioned in the definitions.

Each pupil's protocol was scored by two raters for all the categories. Reliabilities were calculated by intercorrelating the raters' scores for the number of times a particular pupil used a particular category. Since the ratings were to be averaged for further analyses, the Spearman-Brown correction was applied to the correlations between raters. The reliabilities are reported in Table 3.4 which also lists the value categories by their frequency of use.

Correlations among the value categories showed a few scattered correlations which were significant but small. These formed no discernible pattern. The independence between scoring categories was, naturally, partially a function of a quasi-ipsative nature of the data. If the child answered a given response with one value category, he would have used up the answer space and therefore been less likely to respond with another value category. This attenuates positive relationships among the value categories.

The statements which the raters felt would not fit into any of the categories in Table 3.3 were examined along with their suggestions for new categories. Only one new category appeared with sufficient frequency to warrant possible inclusion: Home and family. This category included responses such as "love your parents" and "take care of little brothers and sisters"; it occurred in approximately one out of five subjects' protocols.

As Table 3.4 shows, the reliabilities were generally good. They did range at times down to .5 but these were, generally, for the categories less frequently used. For those categories which were used on the average of at least once per subject, i.e., categories 1 through 7, the reliabilities were acceptable to exceptionally good with the only exception being Diligence.

The following categories were deemed useful for scale development and further analyses: Good Manners, Kindness Toward Others, Academic Achievement, Physical Appearance and Development, Diligence, Obedience, Affiliation, and Honesty. The first seven were included due to their frequency of use and the last was included since it could be a critical value affecting the relationship of the child to socializing agents. If the teacher, for example, saw the child as dishonest, this may be important enough to severely affect the relationship.

Table 3.4

Value Categories: Their Frequency of Use and Reliability

Value Category[a] (listed in order of frequency of use)	Rater Reliability[b] (N = 300)
1. Good manners	.78
2. Kindness towards others by self	.84
3. Academic achievement	.76
4. Physical appearance & development	.94
5. Diligence	.52
6. Obedience	.88
7. Good; unspecified	.62
8. Affiliation	.71
9. Physical harm	.83
10. Mean to others	.53
11. Kindness by others to self	.56
12. Status	.59
13. Honesty	.87
14. Religiosity	.93
15. Imitation	.68

[a]This is not a measure of the pupil's values but only of whether or not he tends to use the value category. Positive and negative usage is scored in the same manner. When the scores were corrected for fluency, the same order occurred except that Religiosity was last.

[b]Reliabilities are based on the Spearman-Brown corrected correlation between two sets of ratings.

The frequency of categories used in this study cannot be directly compared with that of previous research. This project's questions were designed to encourage value comments which would be appropriate for the setting where the research was being conducted. It is obvious that frequency of use of a category is quite dependent upon the type of questions asked. If, for example, another question had been asked such as "what makes a principal admire a child?" the category of Academic Achievement might have moved to the head of the list.

If the research had been conducted in another setting, the total set of categories would probably have been affected. For example, asking a question in the home would have reduced the number of Academic Achievement responses generally produced. Asking the same questions in a church setting would have moved Religiosity nearer to the head of the list. This problem is not deemed critical since there is no set of questions which would not be influenced by the situation or which would be fair to all value categories. Instead, the critical question is whether the categories generated appear to be relevant to the interpersonal relationships in the setting being analyzed, and this appeared to be true here.

Ethnic Differences in Value Categories

Since all of the answers were scored without knowledge of the pupils' SES, sex, or race level, it was possible to analyze the groups for differential usage of the categories. It may be, for example, that one or more of the categories are used exclusively by one subgroup whereas other categories are used in all but one subgroup. If so, measuring instruments would need to reflect such differences.

In searching for differential usage of the value categories by groups of children, it must be remembered that the questions have been designed solely to elicit value categories. They were not designed to tell whether or not the child particularly likes or dislikes any of the values investigated. This particular stage of the research was solely to understand the possible range of values that the child in a particular subculture might naturally use in his discussion. With that understanding, then, it would be possible to talk with the child through questionnaires to determine his values with some assurance that we were speaking the same language.

In the pilot data different children responded with different levels of fluency. In some cases, particularly that of the lower class blacks, the children responded quite fluently giving multiple responses to each question. In other cases, which were more noticeable in the middle class whites, the children were circumspect in their responses. The differences in general fluency could therefore affect the value category usage.

To check for fluency effects in Study I, each child was given

a "total score" which was the sum of all of his responses across all value categories. A three way analysis of variance with race, sex, and SES as the factors was performed on the score. This analysis found no significant interactions but that race and sex main effects were both significant. The means for all main effects are given in Table 3.5. That table shows that whites responded more than blacks and girls responded more than boys. Both lower and middle social class students, with SES defined by fathers' occupational ratings, responded at the same level.

Table 3.5

Group Means for Total Responses in Study I

Race*	Black	16.5	White	18.6
Sex*	Boys	15.4	Girls	19.7
Social Class	Lower	17.8	Middle	17.3

*Significant below .05 level; no interactions were significant.

As a result of the analysis of fluency, the number of times each student used each value category was divided by that student's total score. This controlled for fluency by converting all the value category scores to percentages; the various subgroups could then be compared for their relative use of the value categories.

With sex, race, and SES as the factors, a three way analysis of variance was performed for each value category. The significant effects are summarized in Table 3.6.

It can be noted that there was differential usage of the value categories, particularly for the most popular values. Both Good Manners and Kindness Towards Others had significant three way interactions. The interaction with Good Manners had middle class white males as highest with lower class and middle class black males as lowest. In the Kindness value category, all of the lower class subjects tended to use it more frequently than the middle class subjects except for the lower class black males who tended to use it less frequently.

Value differences were primarily found between boys and girls. Sex formed as many significant relationships as both race and SES combined. Not only were the girls more fluent but they also had a lesser tendency to respond with unspecified comments

Table 3.6

Value Category Usage by Sex,

Race and SES of Pupil

Value Category[a] (listed in order of frequency of use)	Significant ($p<.05$) Relations with Race, Sex or SES Controlling for Fluency[b] (N = 229)
1. Good manners	SES X Race X Sex (McWM&McBF&LcWF > McBM; McBF > LcBM)
2. Kindness toward others by self	SES X Race X Sex (LcBF&LcWM&LcWF > McWM; LcBF > LcBM)
3. Academic achievement	None
4. Physical Appearance and Development	SES (Mc>Lc) & Sex (M>F)
5. Diligence	None
6. Obedience	None
7. Good; unspecified	Sex (M>F)
8. Affiliation	(None; $p<.06$ for race: B>W)
9. Physical harm	SES (Lc>Mc)
10. Mean to others	None
11. Kindness by others to self	Sex (F>M)
12. Status	Race X Sex (WM > BM&BF&WF)
13. Honesty	Sex (F>M)
14. Religiosity	SES X Sex (McM > LcM) Race X Sex (BM > BF)
15. Imitation	Sex (F>M)

[a]This is not a measure of the pupil's values but only of whether or not he tends to use the value category. Positive and negative usage is scored in the same manner.

[b]Abbreviations used: M(male), F(female); B(black), W(white); Lc(lower SES), and Mc(middle SES). All reported differences between individual groups were significant ($p < .05$).

on goodness and were more concerned with interpersonal relationships. The boys tended to be more concerned with personal achievement, for example, physical development. These results are in keeping with common conceptions of sex role differences in the U. S. culture.

Higher socioeconomic status was related to more concern with physical activity and appearance and with less concern for physical aggression. This is in keeping with common conceptions of differences between social classes.

Racial differences occurred only in interaction with SES or sex. These interactions were complex and unique to particular values. For a value such as Status, the white males stand out as different. However, the black males were the unique group on Good Manners. The interactions with Good Manners and Kindness to others mixed race with sex and SES.

Sex, race and SES were not significantly related to some of the classroom related values, i.e., Academic Achievement, Diligence, and Obedience. This can be interpreted as due to the content of phenomena being measured. We were not determining whether the child holds the values but whether the value is salient to the child. It seems that classroom related values are equally salient to most children, a phenomena arising from the similarity of their experiences in the school setting.

Having established differential usage by subgroups on some values, the next question is whether or not the differences of usage are sufficient to condition the usability of the value categories. If, for example, one subgroup does not use one value category, then it might be inappropriate to draw up the scale for that group for that value category. While the mean differences were significant with the sample sizes used here, they did not seem to be major. This is illustrated by Figure 3.1. In this figure, the percent of response in each value category for middle class white females and lower class black males are presented. By picking these two extreme groups, the figure includes both significant interactions and all of the main effect differences attributable to SES, race and sex. It is apparent from this graph that the differences in usages were generally less than 5%. Neither of the groups used a value category which was not used by the other group. It appears that the original decision to use the 7 more popular value categories and Honesty is not inappropriate for either of the extreme groups. Although it may have been more appropriate to select Harm over Affiliation for the lower class black males, the difference between both groups in Figure 1 is not statistically significant for these two values. Both of these groups appeared to partake sufficiently of the same culture so that the same scales can be developed for them.

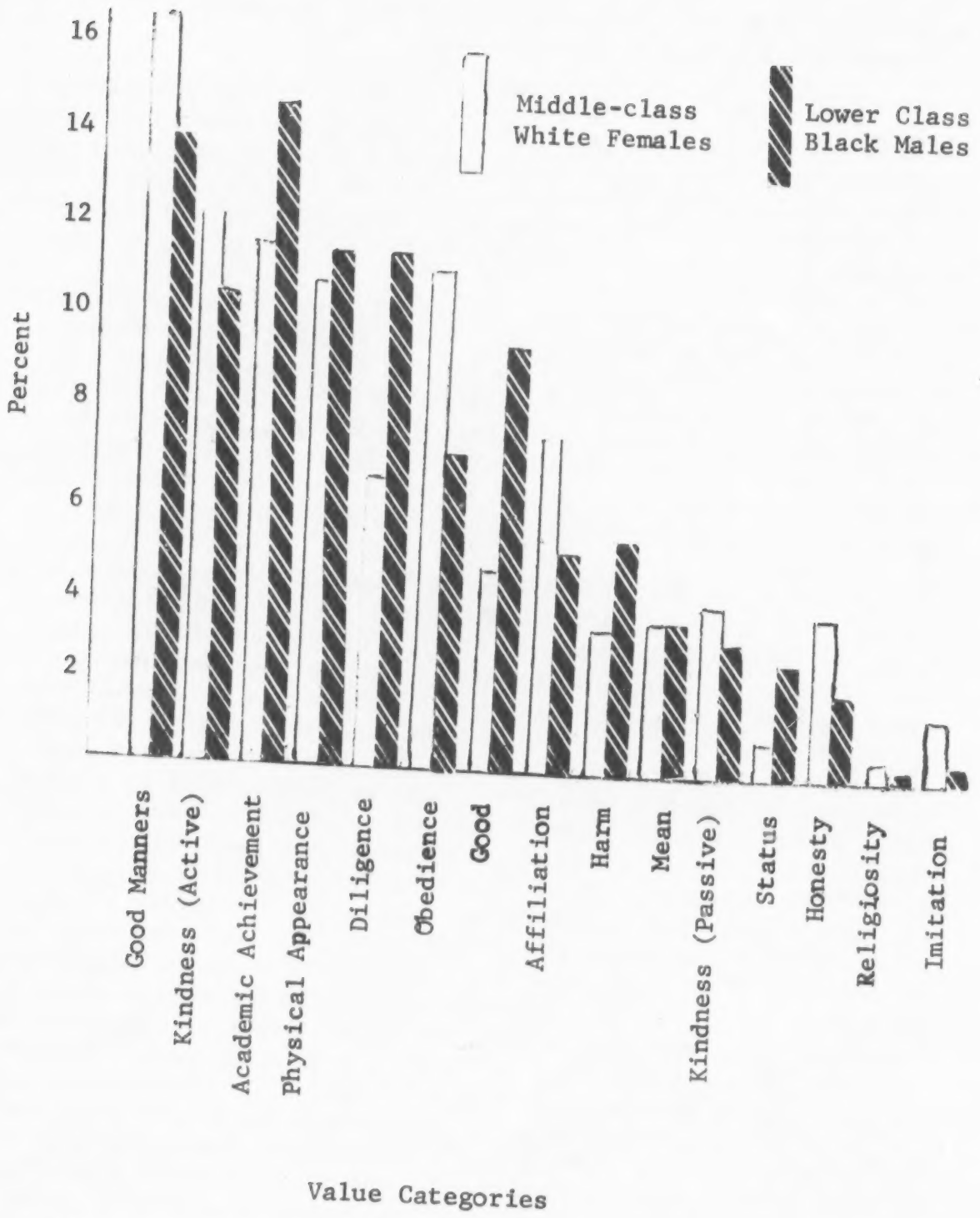


Figure 3.1. Usage of value categories by extreme groups.

CHAPTER IV

MEASUREMENT OF CHILDREN'S VALUES

Measuring Ethnic Differences Fairly

The purpose of developing value questionnaires was to allow ready evaluation of where children stood on each of the values. These individual scores would then be used to determine how values related to the other variables in the project and to examine differences. Since the evaluation of subcultural differences was a major purpose, the value scales would need to be relatively fair to each of the subcultural groups. If, for example, a value scale utilized middle class language, then the resulting mean differences found between lower and middle class children could easily be an artifact of the scale rather than reflecting cultural differences in values. If, on the other hand, the scale were fair to each subculture, then the resulting subcultural differences would reflect the true differences in the dimension of concern.

While the concept of fairness in scales is an obvious one, the operationalization of that concept is not. Indeed, in some senses it is almost impossible. This has proved to be the case in the area of intelligence where someone could argue that any given procedure was biased against one of the subcultures. For example, if a culture fair, non-verbal intelligence test is given to a group of lower class students, they may score lower than middle class students. But is this because they are lower in non-verbal intelligence and the items adequately represent the situation or because their cultural background has not prepared them for this type experience? The question can be answered only if assumptions are made; unfortunately, these assumptions could easily determine what the answer must be.

In the measurement of values, we claim no major breakthroughs in developing culture fair scales. It does, however, seem apparent that there are several principles to be followed to reduce the introduction of one culture's assumptions on another and to increase the fairness of any resulting scales. The first procedure in the area of values has already been described. This involved the use of open-ended questions to determine the value categories and language actually used within each of the subcultures being investigated. Fortunately, that analysis showed that the subcultures being considered in this project do not vary widely in their usage of the value categories. Measuring these value categories therefore means that any resulting scales have a greater probability of being appropriate for the subcultures being studied.

In the construction of questionnaires, several procedures need to be followed to increase the fairness of all of the groups involved.

First, since the values are assumed to be independent of the ability to communicate, the scales should not correlate highly with this ability. Developing only items which have a reading level lower than the grades for which the scales are intended is a beginning in meeting this criterion.

The present project's items were developed to be easier than the fourth grade reading level--as determined by the Spache (1958) formulae-- so that they could be used in fourth and higher grades. A later check on the scale can be made by analyzing its reliability as a function of the verbal ability of the children. If the low-ability children have a usable reliability, then it would appear that the scale is simple enough. If the reliability is independent of the verbal ability level, then it is easy enough for all subjects.

A second procedure to increase the fairness of scales is to evaluate the value items within each group. If, for example, the scale is to predict a given criterion, then each item should predict that criterion equally well within each group. If an item does not function well in any one group, then it would be unfair to that group to use it in the total scale. If the item does function the same in each group, then it can be assumed to be reasonably appropriate for that group.

When a scale is developed where each item is found to be equally related to the criterion within each group, then the selection of items is based upon individual differences within groups. The differences in means between the groups do not enter into the selection or retention of any item. Indeed, these means are independent of the selection process with the only restriction being that no item could be selected on which any group had such a high or low mean that it had little variability. When the resulting scale is then used to compare mean differences across the subcultural groups, there is some assurance that those mean differences are actually mean differences. This bootstrap operation was used in developing the value scales in this study. While it does result in value scales which are reasonably culture fair, there is no assurance that any procedure will create scales which are truly culture fair.

Types of Scales

Several scaling techniques are available that could be used for measuring values. The variety of techniques is large and includes such well known procedures as the Likert, Guttman, Thurstone, and Osgood scales. These scales are all characterized by their summative nature. A person will respond to a number of different items and his responses to those items are summed as his score. Since Osgood's semantic differential was found to correlate about .8 with the Guttman scale and Thurstone scales (Osgood, Suci, and Tannenbaum, 1957), it seems that the various traditional

summative scales are basically in the same measurement category. Since these scales are so widely used, this approach to measuring the value categories was adopted as one procedure for the present project. Due to the cross-cultural stability of Osgood's semantic differential (Snider and Osgood, 1969, Part 5; Osgood, 1964), it would appear to be the easiest to adapt to study subcultural differences.

Scott (1965) has argued that the traditional attitudes approach to values is somewhat limited. It fails to recognize the special nature of values as "oughts" by confusing them with likes and dislikes. The procedure he found useful for measuring values was to describe a particular type of person and ask the subject to indicate how much he admired that person. Thus the subject is not asked if he himself likes that activity but whether a person doing it is admirable or good. This distinction between the admirable and "likes" has some subjective validity since many people feel that they occasionally do what they don't like to do because they ought to. Due to the uniquely value-oriented approach of Scott's scales, it was decided to also include this technique in our attempts to measure the value categories.

Sherif and Sherif (1969) also present a scaling approach which has some theoretical validity for the present project. The procedure calls for having a set of 5 to 10 items which represent different positions on a continuum relative to the attitude or value being measured. The subject then indicates those positions which he finds acceptable, those positions which he cannot accept and that position which is his own. The procedure was theoretically important in the present study because it could provide a measure of the range of acceptance, or tolerance limits, which a teacher or student has for value laden school situations. Unfortunately, we could find no previous case where this procedure had been used with children as young as ours.

Pilot work on several of the scaling procedures deemed potentially useful was carried out independently of the project by Joan Savarese. Her study and findings are reported in Appendix G. In brief, that analysis found little correlation between two methods of measuring values. Because of the impossibility of computing meaningful internal consistency reliabilities for the Sherif scales and the lack of correlations between that scale and the other data collected, it was not known whether it had poor reliability or simply measured something different from the other scales. In addition, the Sherif procedure was also deemed to be somewhat difficult for children of the age tested. Therefore, the Sherif approach was dropped from further consideration.

In all the above scaling procedures, each value is measured directly. All value scores are independent of any other value score in the measurement operations. Rokeach (1968) has argued that a rank order approach is more appropriate. Some of his

analyses (Rokeach, 1971) present evidence that rank ordering value categories does produce meaningful data with older subjects. While the usefulness of this procedure with this age child is not known, a rank order procedure was included in Study III.

For comparison of ethnic differences, a rank order scale has definite limitations. Since rank ordering requires that a person place some values low and some high, no one can rank all the values high or all the values low. And yet it may well be that some children do see all the value categories found in Study I in one direction whereas others evaluate them all in another direction. Such major overall differences would be obscured by the rank order procedure.

A rank order procedure could also magnify minor intergoup differences so that they appeared major. Let us assume, for example, that each group values four given goals quite highly with only very minor differences between them. These minor differences, could lead to consistently different rank orders and therefore the author would conclude that the groups did differ on the values. Whereas the statement that they differed on the values would be true, he would have no way of knowing whether the differences were sufficiently large to be meaningful. Therefore, rank ordering and other ipsative procedures were deemed less important as potential measuring procedures than the non-ipsative scaling techniques. noted above.

Item Development and Administration

Scott Items

Items were developed on the basis of children's responses to the open-ended questionnaire of the first study. Additional items were written by staff members so that there were at least 15 positive and 15 negative items for each of the "major" values and at least 10 negative and 10 positive for each of the "minor" values. The "major" values were those where a good scale was strongly desired. The other value categories were deemed less important.

The prospective items were typed on slips of paper and mixed. Three judges were asked to sort them into value categories, and then pick out two positive and two negative items as the best representatives of the concept being measured. These "best items" and items that were sorted properly by at least two judges were kept in the final stage, giving a total of 24 items for "major values" and 12 for "minor" values. Each item was then tested against a fourth grade reading vocabulary list, and, if necessary, corrected to fit. All items started with the phrase "A child..." and were provided with three answers: "1. I always admire," "2. I sometimes admire," "3. I always dislike." The word "admire" is above fourth grade reading level; therefore it was defined for the children in the following way: " 'Admire' means to think

something is special, to respect someone for doing something, to want to be like someone."

The total of 152 items for this value measurement instrument were then mixed and randomly divided into two halves. There was one restriction on the random assignment. In each half the items were ordered so that each subset of 12 items represented two items for each of the four "major" values and one for each of the four "minor" values. Every set of 12 items was ordered in the same way.

Osgood Items

This measurement technique was used as an alternative approach to Scott. Two different sets of items were included: The first set used one of the Scott items picked as "best" for each of the four scales as the concept to be rated. This set was included to further analyze the relationship between Scott and Osgood approaches. The second set used definitions which were based on three or four of the items used in the Scott instrument for the same value scales.

For each rated value concept or definition, a set of nine bipolar adjective scales was used. These scales were picked on the basis of their belonging to two factors, Evaluation (5 scales) and Potency (4 scales), based on prior studies (Osgood, Suci & Tannenbaum, 1957). It was felt that only these two factors are relevant and important in the measurement of values.

Administration: Study II

The administrators were all young women. Race of the testers was counterbalanced across the SES and race of the classrooms. All items were read aloud by testers. The Osgood Scale was administered in between the two parts of the Scott Scale.

Similarity of Scott and Osgood Scales

In Study II, we attempted to measure the same constructs by Scott and Osgood scales as a check on their overlap. If extensive overlap was found, then it would not be necessary to include both sets of scales. If little overlap was found, then both scaling procedures would be included in Study III.

Four identical items were included among those in the Scott scale and those in the Osgood scale. Each item was worded the same except, as required by the Scott approach, the stem of "a child" was used to begin each item there. The four separate items represented four different value categories. The correlations between the individual Scott and Osgood items are presented in Table 4.1 for the groups in which the correlation was computed. An examination of this table shows that the correlations were not spectacularly high. While the test-retest of any individual item is

Table 4.1

Correlations Between Scott and Osgood Items

Items	Correlations Between Scott Item and Semantic Differential Evaluation Rating of the Item			
	Middle Class		Lower Class	
	Blacks N = 97	Whites N = 99	Blacks N = 199	Whites N = 127
(A child)[a] saying "Yes, Sir, No, Sir"	.53	.48	.19	.46
(A child) thinking of others' feelings	.27	.27	.30	.39
(A child) belonging to a group	.46	.47	.31	.46
(A child) liking to do extra work for school	.52	.58	.27	.50

[a]Used only in the Scott form

never perfect, such correlations would be expected to vary from .5 to .8 for the items to be considered equivalent. Since the highest correlations in this case were about .5, it does appear that these two procedures are measuring different aspects of the subjects' reaction to the values. It was therefore decided to develop scales for measuring the value categories by both of these procedures.

Correlations Between A Priori Scales

The first step in analyzing the data collected on the item pool was to score it for the a prior scales. These scales were formed out of the items which all judges were able to place in the value category for which the item had been constructed. The internal consistency reliability (coefficient alpha) for each of the a priori scales was calculated. The correlations among the scales were also computed.

It was felt that an examination of the Scott a priori scales' reliabilities and correlations would provide several kinds of basic information. If the reliabilities for one or more scales were exceptionally low, then there would be little hope for developing a scale for that category. Reasonable reliabilities would suggest

that we had placed together items which the children themselves also saw as going together.

Even though the values had appeared equally in all groups, there was a possibility that some groups might differentiate between them whereas others might collapse the values. To provide a check on this possibility, the reliabilities and correlations were calculated separately in each SES by race group.

The correlations among the Scott a priori scales could lead to several different conclusions, assuming the reliabilities to be adequate. If these correlations were extremely low for each group, then the project would need to further develop a separate scale for each value category. However, it was a possibility that the correlations between the scales would show the same pattern of high correlations and low correlations across the groups. If that appeared to be the case, then further evaluations would be warranted to determine whether or not to collapse certain of the scales. Differential group patterns might lead to keeping all of the value scales to allow for the complex patterning.

Instead of the data following any of the options suggested above, the correlations between the scales were all systematically quite high for all groups.

Table 4.2 presents the reliabilities and correlations among the a priori value scales measured by the Scott approach for one of the subgroups (it is essentially the same as the matrices for the other groups). In the upper half of the matrix are the raw coefficients. The diagonal contains the reliabilities and the lower half of the matrix gives the correlations among the scales corrected for their measurement errors. This lower triangle is, therefore, the theoretical relationship between the constructs being measured by the scales.

This matrix shows the correlations to be all systematically quite high relative to the reliabilities. Indeed, the matrix suggests, for example, that the Diligence scale is as good a measure of Honesty as it is of Diligence, regardless of what the content of the items may be. With only one or two exceptions, each and every scale correlated so highly with the other scales as to be a parallel form of those scales. Therefore the conclusion to be drawn from this matrix is that there is one general factor underlying all of the scales. While other minor factors might be found if a formal factor analysis were computed, such factors could only be trivial in size. We have chosen to call this general factor VALUE SOCIALIZATION.

Several of Osgood scales were also included in this data collection. They were scored and intercorrelated in the same manner as the Scott scales with the results being given in Table 4.3

Table 4.2

Reliabilities of and Correlations

Among Scott A Priori Scales:

Lower-Class Blacks

		1	2	3	4	5	6	7	8
1.	Academic Achievement	.86	.64	.83	.85	.84	.83	.82	.64
2.	Affliliation	.91	.57	.66	.64	.66	.70	.62	.57
3.	Diligence	1.02	1.00	.77	.83	.86	.85	.84	.63
4.	Good Manners	.98	.90	1.01	.88	.87	.88	.91	.67
5.	Honesty	1.01	.97	1.08	1.02	.82	.86	.85	.72
6.	Kindness	.96	.99	1.03	1.00	1.02	.87	.86	.67
7.	Obedience	.96	.88	1.03	1.05	1.01	1.00	.85	.62
8.	Physical Appearance	.86	.95	.90	.90	.86	.89	.84	.64

Note: The internal consistencies are in the diagonals, the raw correlations above the diagonals, and the correlations corrected for attenuation below the diagonal. Values over 1 are possibly due to sampling error. N=199.

for one of the groups.

Two forms of the semantic differential scales were available, Evaluation and Potency. Internal consistency reliabilities were computed for each scale. In this case, such reliabilities do not indicate the degree of generalizability across variations in the constructs but in variations in the adjectives used to measure the Evaluation or Potency factors. The only meaningful corrections for attentuation that can be made are of the correlations between the two scales used for rating a particular concept. These corrections are given below the diagonal and suggest that, in the realm of values, distinguishing between Evaluation and Potency is difficult. In Study III, the two factors were combined; a person scored high on a value only if he saw it as both highly evaluated and potent.

The correlations among the values are much the same for the semantic differential scales as for the Scott a priori scales except not quite so extreme. A general factor also appears to

Table 4.3

Reliabilities of Adjective Scales and Correlations Among Semantic Differential Measures of Values: Middle-Class Whites

	1	2	3	4	5	6	7	8
Academic Achievement								
1. Evaluation	.86	.78	.48	.47	.40	.42	.50	.42
2. Potency	.91	.85	.47	.59	.29	.46	.44	.60
Affiliation								
3. Evaluation			.84	.72	.39	.36	.60	.53
4. Potency			.87	.84	.23	.50	.48	.62
Good Manners								
5. Evaluation					.92	.81	.52	.42
6. Potency					.90	.89	.44	.59
Kindness								
7. Evaluation							.83	.73
8. Potency							.87	.86

Note: The internal consistencies are in the diagonals, the raw correlations above the diagonals, and the correlations corrected for attenuation below the diagonals. N=99.

occur but is less prominant than with the Scott scales. It appears that a measure of the general factor, such as summing across all the values, would be a legitimate addition to the semantic differential scores.

Item Factor Analyses

While the data given in Table 4.2 show a general factor between those scales, such a general factor could occur because the scales did not adequately represent the value differences which were inherent in the items themselves. If this were the case, each scale would partake of the same set of values and could correlate together at a fairly high level as parallel forms. Whereas the reliabilities are a bit high to suggest that this is a powerful effect, it is a possible effect. Therefore factor analyses of the Scott items were pursued to determine if any non-hypothesized value distinctions could be identified.

The item pool was not especially designed for an exploratory factor analysis. Since the items were written to embody certain

concepts, the most appropriate type of factor analysis would involve multiple group or maximum likelihood hypothesis testing. However, the results of such an analysis are essentially known from the reliabilities and correlations. If the factors were placed through the hypothesized or a priori scales, they would correlate too highly to warrant further consideration. This conclusion is reached from Table 4.2 which presents what are approximations to the correlations between such multiple group factors. Therefore an exploratory principal axis analysis with analytic simple structure rotation appeared to have the greatest potential of providing new information.

As the number of items increases, the possibilities of capitalizing upon chance in a factor analysis increase and therefore the number of subjects should likewise increase. The item pool, totaling 152 items, was large relative to the number of subjects. Therefore the resulting factors would be more likely to be influenced by chance than is advisable. In addition, this large number of items means that most programs for factor analyzing data cannot be run.

To solve the problem of the unstable factors due to large numbers of items and to be able to use local computer programs, the item pool was split in half. Each of these random halves was factored on the same subjects and the results compared. If one or more factors were warranted by the data, then they should appear in both analyses. Any factor which appeared only in one analysis would be likely to have arisen by chance or to be trivial in overall importance.

The number of factors indicated by the commonly used Guttman's criterion were extracted and rotated to the Varimax-Promax positions (Harman, 1967; Gorsuch, 1970). This procedure does not allow a general factor (Gorsuch, 1966) but does search for simple structure factors. The four items with the highest correlations with a given factor were selected and are given in Table 4.4, provided that all such items had a correlation greater than an absolute value of .4 with the factor.

Table 4.4 contains the salient items from both analyses. By comparing the item content across the two factor analyses, it can be seen that no interpretable value oriented factor replicated across the sampling of items.

It now appears that children's values are relatively undifferentiated. While fourth and fifth grade children do use value categories which adults see as distinctively different, they tend to equally hold or not hold to all of these categories at the same time. They group together behaviors such as being kind towards others, good manners, academic achievement, diligence, and obedience into one broad value socialization dimension.

Table 4.4

Factor Loadings of Scott Items (.40 and Above)

on Two Randomly Divided Halves of Questionnaire

First Half		Second Half	
Scale	Factor A	Scale	Factor A
Hon.	Telling the truth .74	A.A.	A child getting good grades .72
P.A.	Being clean .71	Dil.	Doing a good job, even if it takes a long time .70
G.M.	Being quiet when the teacher is talking .71	Dil.	Coming to school every day .69
Ob.	Obeying school rules .68	G.M.	Saying "Good Morning" to the teacher .68
	Factor B		Factor B
G.M.	Talking back to the teacher .49	Kind.	Laughing at someone's worn out clothes .65
G.M.	Interrupting the teacher .48	G.M.	Talking back to his parents .62
A.A.	Giving wrong answers on a test .46	Ob.	Doing what he wants instead of listening to people who know better .60
Aff.	Working alone instead of with a group .42	A.A.	Playing rather than studying for a test .60
	Factor C		Factor C
A.A.	Not paying attention to lessons and books that are difficult .40	Kind.	Telling stories about other kids .52
Aff.	Telling another child's secret .40	Kind.	Laughing when someone is hurt .52
Hon.	Being able to admit he did something wrong .40	G.M.	Interrupting someone else who is talking too much .49
	Factor D	Kind.	Making others happy .47
Kind.	Trying to get somebody into trouble .72		Factor E
Ob.	Doing things the teacher has said not to .70	G.M.	Talking about someone behind his back .62
Kind.	Hurting other children's feelings .68	Hon.	Lying to the teacher about forgetting homework .57
Ob.	Breaking school rules .68	G.M.	Interrupting what someone else is saying .53
	Factor E	Kind.	Calling another child bad names .51
Ob.	Not doing what he is told as long as he can get away with it .59		Factor F
Hon.	Cheating on a test .56	A.A.	Wanting to learn something new every day .45
Dil.	Doing a sloppy job .52	G.M.	Raising his hand before talking .43
Kind.	Walking past another child who is hurt .51		Factor L
	Factor F	G.M.	Letting other children have their turn .55
Ob.	Doing what the teacher says .66	Ob.	Doing what parents say
Kind.	Helping everybody who needs help .60	Aff.	Liking to be with people

First Half

Kind. Helping children who are sick .56
A.A. Studying hard to become best student in class .51

Factor H

Kind. Breaking up a fight .47
G.M. Saying, "Yes, Sir,"and "No, Sir" .43
G.M. Being quiet when the teacher is talking .42
A.A. Finishing homework even if it means missing a TV show .41

Factor I

A.A. Not caring if he becomes well educated or not .50
G.M. Talking back to the teacher .42
G.M. Interrupting the teacher .42
Kind. Making fun of another child who has braces on his teeth .41

Factor M

A.A. Studying hard to become the best student in class .54
G.M. Acting right .53
G.M. Cussing .50
G.M. Saying, "Yes, Sir " and "No, Sir" .49

Second Half

Dil. Doing a good job even if it takes a long time .48

Factor N

A.A. A child hating to study .60
A.A. Feeling that studies get in the way of having fun .54
Ob. Doing what he wants instead of what his parents ask .52
Dil. Leaving the job for someone else to finish .48

Factor O

A.A. Doing homework before doing other things .49
G.M. Getting out of the classroom quietly when dismissed .45
Ob. Following directions as they are given .45

Abbreviations used:

A.A. Academic Achievement
Aff. Affiliation
Dil. Diligence
G.M. Good Manners
Hon. Honesty
Kind. Kindness
Ob. Obedience
P.A. Physical Appearance & Development

The Resulting Scales

Based on the above considerations, it was decided to build one value scale from the Scott items. This would represent the general thrust of the value concerns of children of this age.

The item analyses were carried out within each of the eight groups formed by the categories of SES, sex, and race. Within each of these eight groups, each item was correlated with a total score off the entire item pool and with the child's grade. The former was used to evaluate how well the item measured the central value socialization thrust of the children's responses. The latter was used as a rough indication of the degree to which the item might be influenced by abilities. This follows since the older children are somewhat brighter than the younger children. Any item which correlated significantly with grade in any of the eight groups was

eliminated on this score. Requiring the item to correlate significantly with the total item pool meant that, naturally, the item had to be reasonably understandable by each group of children and measure the central thrust of the item pool. The minimum correlation necessary to reach statistical significance varied from .24 to .18, depending on the N in each group. The minimum correlation accepted was .23, the highest was .77, and most were in the .4 to .65 range.

Note that the item-total criterion assumes that the thrust of the total pool is the same in each of the eight groups. Inasmuch as this is so, this item selection procedure retains only those items which are relatively fair to all the groups involved. This assumption is reasonable in the present data since the prior analysis of naturally occurring value categories suggested that the children did indeed talk in these terms and the subgroup correlations between a priori scales suggested that the same general factor occurred. The fact that many of the items were written from the actual children's response regardless of whether the child was black or white, male or female, or middle or lower class likewise suggests that the assumption would be relatively accurate. But the assumption of the equivalence of the total pool is one of the reasons a completely culturally fair scale is difficult to develop.

Out of the 152 items examined, only nine met the relatively strict multiple criteria proposed here. The resulting Value Socialization Scale was almost counterbalanced for response set with five positive answers and four negative answers being scored as high social conformity. It included items from five of the eight original values and no value had more than two items included. These items are given in Table 4.5 along with the median of their

Table 4.5

Value Socialization Items

Item	High Scoring Response	Median Item-Total Correlation
A child doing what he wants instead of what his parents ask.	Admire	.58
A child doing the best he can.	Admire	.73
A child getting by in school with little work.	Dislike	.35
A child giving wrong answers on a test.	Dislike	.36
A child doing what the teacher wants.	Admire	.46
A child talking back to his parents.	Dislike	.54
A child being kind to others.	Admire	.48
A child talking back to the teacher.	Dislike	.56
A child going to the store for somebody.	Admire	.46

item total correlations. This median was computed by averaging the Fisher Z's of the two correlations in the middle of the range formed by the eight correlations. Note that a person of any age can respond meaningfully to these items but that his response would center on children.

An additional item was added to the nine for Study III. This tenth item was used as a warm up, non-scored item. (Appendix D contains the scale as given to the children and teachers in Study III).

Since the semantic differential adjectives had been selected on the basis of past research, item analyses were less critical than with the Value Socialization Scale. When individual item analyses were run within the four groups formed by crossing SES and race, the item-total correlations were reasonable. The best items were selected from among the nine original ones and are given in Table 4.6. Any person of any age can respond meaningfully to these adjectives.

Table 4.6

Semantic Differential Adjectives

Evaluation	Potency
Happy - Sad	Strong - Weak
Nice - Awful	Sharp - Dull
Good - Bad	Big - Small

The concepts for the semantic differential value scales were developed on an a priori basis. Since it was desired to cut across all of the eight values to estimate the total socialization effect, each of the eight values was rated. The definitions given earlier were simplified somewhat for use with children with samples from the Scott items being included when those items correlated well with the total score from that value's a priori Scott style scale. These value definitions are given in Table 4.7 (the instrument as given in Study III is in Appendix D). Note that including all eight values means that analyses could proceed in two different ways. On the one hand, the total score across the values can be calculated. On the other hand, each of the values can be independently scored as desired.

To increase the flexibility of the approach and to possibly look at individual values where they do differ, one further scale was developed on an a priori basis. This required the rank ordering of the eight values by the children. Preliminary trials suggested

Table 4.7

Value Concepts for Semantic Differential Rating

and Rank Ordering

(A CHILD'S)[a] BEING KIND TO OTHERS (for example, thinking of others' feelings, breaking up fights, making others happy).

(A CHILD'S) OBEYING PEOPLE WHO KNOW BETTER (for example, doing what the teacher tells (him) you, obeying school rules, stopping play to do what you are (he is) told).

(A CHILD'S) BEING HONEST (for example, bringing back the correct change from the store, being able to admit you (he) did something wrong, keeping your (his) promise).

(A CHILD'S) BEING WELL LIKED BY OTHERS (for example, sharing special secrets, having a lot of friends, being liked by other children).

(A CHILD'S) BEING GOOD LOOKING (for example, being clean, spending time taking care of the way you (he) look(s), getting a good grade in Physical Education).

(A CHILD'S) DOING WELL IN SCHOOL WORK (for example, studying hard to become one of the better students in class, finishing your (his) homework even if it means missing a TV show, liking to do extra work for school).

(A CHILD'S) TRYING TO ALWAYS DO YOUR (HIS) BEST (for example, finishing every job you (he) start(s), giving up playing when you have (he has) a task to do, working on your (his) job even when it is dull).

(A CHILD'S) HAVING GOOD MANNERS (for example, saying "good morning" to the teacher, getting out of the classroom quietly when dismissed, waiting in line for your (his) turn).

[a]"A child's" was the phraseology used for the Rank Order Values Scale.

this was possible. The values as defined for the children were the same as those used in the semantic differential. (This instrument is also included in Appendix D.)

Characteristics of the Scales in Study III

The value scales developed in this project were used in the third study. Study III also included several other measures of

general personality characteristics. While the prime analyses of these data are reported in later chapters, part of the data does reflect upon the adequacy of the scales.

The intercorrelations among the Osgood evaluation and potency versions of the same scale were high, as was also found in Study II. For this reason, all analyses combined the evaluation and potency scales into one general score.

Within each of the eight sex by class by race groups, internal consistency reliabilities in Study III were calculated for the scales. These are given in Table 4.8. For comparison

Table 4.8

Reliabilities of the Value Scales
Within SES x Sex x Race Subgroups

	Middle Class				Lower Class			
	White		Black		White		Black	
Internal Consistency Reliabilities (Alpha)	Female	Male	Female	Male	Female	Male	Female	Male
A. Value Socialization Scale								
Item Selection Sample:								
Study II	.73	.81	.70	.75	.64	.81	.78	.78
(No. of Subjects)	(47)	(52)	(50)	(47)	(58)	(59)	(114)	(85)
Cross-Validation Sample:								
Study III	.75	.73	.79	.76	.77	.63	.72	.72
(No. of Subjects)[a]	(90)	(74)	(55)	(50)	(29)	(21)	(65)	(41)
B. Semantic Differential:								
Study III								
Total Value Reliability[b]	.91	.88	.94	.93	.87	.93	.90	.92
Value plus Adjective Reliability[c]	.96	.95	.96	.95	.94	.95	.94	.95
(No. of Subjects)[a]	(89)	(74)	(55)	(51)	(28)	(22)	(65)	(41)

[a]N's vary slightly in Study III due to missing data for some subjects.

[b]Calculated by applying coefficient alpha formula to variances of each value and total variance; it measures reliability across value constructs.

[c]Calculated across both values and adjectives; this is the usual way of computing semantic differential reliabilities.

purposes, the Value Socialization Scales reliabilities from Study II are also included (the Semantic Differential of Study II was not comparable.) There is no discernable trend for any of the groups to have distinctly poorer reliabilities than any other of the groups. For example, the group with the lowest reliability in Study III had tied for the highest reliability in Study II. These reliabilities are all adequate for research purposes and the scales seemed appropriate for each group. The reliabilities of the Value Socialization scale were also calculated as a function of the children's verbal ability. An analysis of these reliabilities (cf. Appendix F) suggested that the reliability was adequate for the lowest verbal ability children. No relationship was found between ability and this scale's reliability.

Several sets of correlations with the value scales could be computed. The first correlations of interest are those among the value scales; these are given in Table 4.9. Once again it is apparent that the semantic differential approach to measuring one's feelings about values differs from the approach used in the Value Socialization Scale which attempts to measure the value itself. These do appear to be somewhat distinctive domains. In like manner, the rank order procedure does not correlate particularly well with the Semantic Differential Value Scales. This may be partially a function of the ipsative nature of the Rank Order Value Scale, of independence between the scales, or of low reliability within the Rank Order Value Scale. Despite the pretests, rank ordering eight values appeared to be somewhat difficult for the children and we are not confident of the result.

The second set of correlations is with the other measures on the children. An examination of Table 4.10 shows that a correlation was found between the Value Socialization Scale, the Semantic Differential Scales and intelligence. Perhaps this is a function of the fact that intelligence generally correlates positively with most socially desirable characteristics. Perhaps it is a function of the fact that brighter children learn faster in the academic setting and therefore learn the value structure faster.

It is reassuring to note that the value scales did not generally correlate with the Social Desirability scale. Further checks were made to assure the independence of the Value Socialization and Semantic Differential Total Value scales from social desirability response set distortions. Since the correlations could have been low due for several reasons that would have obscured a real relationship (e.g., a lack of homoscedasticity), the means and variances of the highest third of the scorers on the Children's Social Desirability scale were compared against those of the lowest third on both the value scales. None of these analyses suggested any distortion effect except, possibly, that the higher third tended to avoid extreme responses on the value scales (they probably did

Table 4.9

Correlations Among Value Scales

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	Semantic Differential																		
1.	Academic Achievement	1.00																	
2.	Affiliation	.59	1.00																
3.	Obedience	.54	.57	1.00															
4.	Good Manners	.58	.63	.64	1.00														
5.	Honesty	.48	.51	.52	.64	1.00													
6.	Kindness	.51	.57	.52	.65	.65	1.00												
7.	Physical Development	.49	.50	.46	.60	.62	.58	1.00											
8.	Diligence	.51	.50	.53	.62	.58	.63	.64	1.00										
9.	Total	.74	.78	.77	.85	.79	.81	.77	.79	1.00									
10.	Socialization	.22	.20	.22	.28	.30	.27	.19	.20	.30	1.00								
	Rank Order																		
11.	Kindness	-.00	.01	-.05	-.03	-.01	-.00	-.04	-.08	-.03	-.01	1.00							
12.	Obedience	.04	-.02	-.02	-.02	-.03	.02	-.08	-.05	-.02	.05	.07	1.00						
13.	Honesty	.03	-.01	-.01	-.02	-.03	-.03	.06	.05	.00	-.14	.03	-.03	1.00					
14.	Affiliation	.05	.03	-.02	.04	.02	.09	.04	.05	.04	.19	-.11	-.02	-.19	1.00				
15.	Physical Development	.07	.02	.02	.03	.13	.08	.04	.06	.07	.15	-.16	-.26	-.27	.06	1.00			
16.	Academic Achievement	-.00	.07	.07	.10	.15	.10	.13	.07	.11	.03	-.35	-.19	-.22	-.15	-.06	1.00		
17.	Diligence	-.10	-.07	-.01	-.04	-.15	-.10	.00	-.01	-.07	-.17	-.23	-.25	-.13	-.32	-.17	-.02	1.00	
18.	Good Manners	-.08	-.03	.01	-.06	-.07	-.13	-.13	-.07	-.09	-.08	-.37	-.29	-.20	-.19	-.15	.01	.11	1.00

Note: N = 428; .05 significance level = .10; .01 significance level = .13.

Table 4.10

Correlations of Value Scales
with Personality Characteristics

	Personality Variables					
Semantic Differential	Culture Fair Intelligence Test	Verbal Analogies	Children's Social Desirability	Extroversion	Anxiety	Teacher Potency
Academic Achievement	.07	.13	.13	.14	-.08	.32
Affiliation	.14	.20	-.01	.16	.00	.36
Obedience	.13	.18	.06	.10	.00	.45
Good Manners	.13	.21	.08	.16	-.03	.46
Honesty	.19	.29	-.03	.18	-.11	.39
Kindness	.17	.23	.02	.12	-.02	.40
Physical Development	.09	.17	.05	.12	.01	.40
Diligence	.02	.12	.13	.10	.00	.50
Total	.15	.24	.07	.17	-.03	.52
Value Socialization	.38	.41	.00	.11	.16	.22
Rank Order						
Kindness	.08	-.06	-.03	.03	-.01	-.05
Obedience	-.04	.03	.04	-.03	-.10	-.04
Honesty	.21	-.20	.05	-.01	-.00	.01
Affiliation	.22	.14	.07	.00	-.11	-.02
Physical Development	.11	.14	-.00	.05	.00	.03
Academic Achievement	.06	.12	-.12	.00	.06	.08
Diligence	.22	-.17	.08	-.06	.13	.03
Good Manners	.06	.00	-.07	.00	.03	-.04

Note: N = 428; .05 significance level = .10;
.01 significance level = .13.

not want to be seen either as too good or as bad, but rather as "regular guys and gals."). Nevertheless, the possible alternative hypothesis that some of the high scorers were high because they were responding in the socially desirable direction was kept in mind during the data analyses. On three occasions, the high value scorers were split into upper third and lower two-third groups on the social desirability scale to see if their value scores functioned differently in prediction situations (cf. Chapter 7 and Appendix H). Both the Value Socialization and Semantic Differential Total Value scales seem undistorted by that which leads to high scores on the Children's Social Desirability scale. We can only assume that the children have been reasonably honest.

The correlations between teacher potency and the semantic differential scales were higher than expected and may reflect a method bias: both were measured by the same semantic differential technique.

The correlations between the different concepts measured by the Semantic Differential Value Scales are similar to those given in Table 4.3 earlier. This again suggests that a total score from the Semantic Differential Scales is a meaningful concept although the values can be distinguished somewhat. Whether or not these separate values lead to differential prediction is a matter for further research.

Content and Meaning of the Scales

Both the value categories that children use and the items which were empirically good measures of children's values contain elements of obedience and conformity. It could be that the value categories and items were partially a function of doing the research in traditional school classroom settings. Here the child must sit in his chair and relate to the one teacher who is the obvious authority in the classroom. A teacher can only survive in such an atmosphere if she does instill some respect for authority, obedience, and other such characteristics in the children.

Supportive evidence for a general socialization factor emphasizing conformity is found in the psychological literature. In several studies with young adults and children, Cattell (1957) found a factor labeled G: Superego. G is the factor which Gorsuch & Cattell (1967) found to form a second order personality factor independent from other personality factors such as anxiety, extroversion and intelligence. Sells' (1968) analyses comparing Cattell's and J. P. Guilford's work suggests this social conscientiousness factor is a prime motif in Guilford's personality questionnaires as well. The present scales' content and independence of personality suggests it may be the same as Cattell's G and Sell's social conscientiousness factor.

Other studies also suggest the pervasiveness of a social conformity dimension. J. S. Guilford (1971) analyzed values from a broad definition, i.e., analyzed likes and dislikes. In working with slightly younger children than the present study, she found a strong factor analytic dimension which was called social conformity. In content, this factor closely resembles the present one. It therefore appears that there is widespread evidence for such a factor in both children and adult research.

Another line of thought would relate this factor to stages of moral development. Kohlberg (1969) has identified such stages and noted that children pass through a social conformity level. This is characterized by some concern for others at the first stage of the conforming level and a legalistic approach at the second. Such children would score quite high on the present scale. Kohlberg holds that children must pass through a conventional socialization in order to be able to transcend that orientation and move into a principled concern for justice which involves love for others.

Future research is definitely needed to clarify the meaning of this dimension which has been so widely found. Not only is its relationship to Kohlberg's stages in doubt, but the lack of a correlation between our two forms of measuring it is also an unsolved problem.

CHAPTER V

VALUE DISCREPANCIES AS A FUNCTION OF SOCIAL CLASS, SEX AND RACE

Introduction

A discrepancy in values exists when the teacher has a different value or values than the student. Value discrepancy can arise from multiple sources, one of which is the role the society holds for that child. This role is primarily the function of such variables as social class, sex, and race. This chapter analyzes the extent to which differences in values are associated with these role related variables.

Since a discrepancy measure is found by simply subtracting the pupil's value score from the teacher's value score, a large value discrepancy associated with any of the three background variables can arise because the teachers in one type of school have different sets of values than do teachers in other schools, or because the students there have different sets of values, or because both teachers and pupils have values which differ by SES, sex, or race. Therefore the statistical analyses will be of the teacher and pupil scores individually. When conclusions are reached as to the relationship of these values to the SES, sex, and race variables, their relationship to discrepancy scores will also be apparent.

The Relationship of the Teachers' Values to the SES and Racial Composition of Their Schools

In the first testing session of Study III, the teachers and students responded to the same value instruments. The teachers' scores were then entered into analyses of variance to determine whether teachers at the various kinds of schools have the same mean values. The classroom grade was not entered into these analyses since it was felt to be less important and since including it would have halved the already limited number of subjects per cell.

For each value scale, a two way analysis of variance was computed. The two factors were the race of the students in the school and the school's socioeconomic status reputation. The results are summarized in Table 5.1. This table shows only a few significant effects with most of those effects being main effects rather than interactions. In the Semantic Differential scales, the teachers in the lower class schools were more concerned with academic achievement than those in the middle class schools. The teachers in the white schools were more concerned with kindness than those in the black schools. On the total score, teachers in white schools gave higher ratings to the values than did the teachers in the black schools. Several similar differences can

Table 5.1

Summary of Analyses of Teachers' Values
by School Race and School SES

Scale	Results*
Semantic Differential	
Academic Achievement	Lower class school teachers higher than middle class school teachers
Affiliation	No difference
Diligence	No difference
Good Manners	No difference
Honesty	No difference
Kindness	White schools' teachers higher than black schools' teachers
Obedience	No difference
Physical Appearance & Development	No difference
Total Score	White schools' teachers higher than black schools' teachers
Socialization	No difference
Rank Order	
Academic Achievement	No difference
Affiliation	No difference
Diligence	Black schools' teachers higher than white schools' teachers
Good Manners	No difference
Honesty	No difference
Kindness	Black lower class school teachers higher than others
Obedience	No difference
Physical Appearance & Development	No difference

*All differences reported were significant at the .05 level or better by two way analyses of variance.

be noted in the rank order scale although the socialization scale had no significant relations with SES or race of schools.

It is interesting that the teachers in white schools felt more positively towards the values included in the Semantic Differential than did those in the black schools. As will be noted later in this chapter, the differences between these two sets of teachers is as great as any of the differences found on the same scale among the children. This difference also produced some interesting results reported in the next chapter.

Ready explanations for the higher scores of teachers in white schools are many. But in drawing up such interpretations, it should first be noted that teachers in black schools are predominantly white. Only two out of the eight teachers in the black schools were black (all the teachers in the white schools were white). Second, the difference cannot be attributed to differences in experience within the schools for the teachers in the white schools had a mean of 15 years experience and those in the black schools had a mean of 12 years of experience, a difference not statistically significant.

Probably the explanation of the differences between the teachers must take into account one fact: the two sets of teachers were found to be significantly different in length of time at their schools. Teachers in the white schools were, on the average, in their ninth year at that school whereas those in the black school were, on the average, in their second year. This latter statistic is a bit misleading due to the skewness of the scores. Actually, five of the teachers in the black schools were there for the first year whereas none of the teachers in the white schools were in their first year at that school.

This situation probably arose due to the integration of teachers that took place immediately before schools opened for the year during which this data was collected. At that time, principals were allowed to select out some teachers who would not be shifted and the rest of the teachers were placed into a lottery. These teachers were then randomly assigned to schools with the only restriction being that the proper ratio of black to white teachers appeared in each school. The end effect was to replace most of the teachers in black schools with white teachers whom the principals had not identified as their best. But in the white schools, due to their greater number, a much smaller percent of the teachers were replaced.

There may also have been a teacher selection of those who participated in the study. Most of the schools tested have multiple fourth and fifth grades and the principal recommended one of his teachers to participate. The reasons for the recommendation are not known and some of the differences identified could be correlated

with differential conceptions of what the research would be like or other such variables which could be associated with race of the school.

Another possible explanation is that the experiences of the teachers in the school during their first several weeks had altered their values. But if this explanation were true, then it would be difficult to understand why their values did not change more over the semester (cf. Chapter VI).

The reasons suggested above for why teachers with certain value structures were found more often in some schools compared to others are quite tentative. Others could also be proposed. At the present time, we feel that our best comment on the ultimate cause of this difference is "we don't know."

Social Class, Sex and Race Differences in Children's Values

The State of the Literature

It is commonly assumed that the values held by blacks differ from those held by whites and that lower class children hold to different values than middle class children. This is true of some of the educational literature. For example, Cloward and Jones (1963) believe that the difficulties black children encounter in school are attributable to the failure of the school to recognize their distinctive system of values. Cheyney (1966) supports this belief by holding that school teachers generally come from middle class homes whose values conflict with those of disadvantaged children.

What evidence is there that value differences between sub-cultural groups occur? Reviews of the research literature such as that provided by Guilford (Guilford, et. al.,1971, pp. 8ff.) note that there are a number of studies where social class differences have been found. But the results from these studies are neither clear nor consistent with one another.

In regards to school related values such as academic achievement, it has been concluded that middle class children are higher than lower class children (Hyman, 1963; Riesman, 1962; Rosenberg, 1965). Other studies, however, found no evidence of differences between lower and middle class children (Godbold, 1967; Brozovich, 1966; Bradford, 1967). Still other studies (e.g., Baker, 1967; Kohn, 1969) can be cited which found lower class children valuing academic achievement more highly than middle class children.

The same confusion exists in analyses of racial differences. Evidence has been reported that whites value academic achievement

more than the blacks (e.g., Pettigrew, 1964; Rosen, 1959) while others (e.g. Summerskill, 1962) support the other side.

In addition to the above positions, still another one exists. It can be argued that despite any similarities or differences between the values of different subgroups, there is differential perception of the opportunity to achieve even those values and goals which are held in common. Sherif and Sherif (1964) were led to this conclusion from their study of adolescents as were Landis, Dintz, & Rickloss, (1969) when they studied the differential perception of the life chances of sixth grade boys.

Because such studies are not consistent in their findings, meaningful integrated reviews of studies of ethnic and social class differences would be extremely difficult if not premature.

Problems in Analyzing Ethnic Differences

But not only are the results of these studies inconsistent, they are limited in their meaningfulness. Obviously many differences between subcultural groups in physical features, attitudes , life styles, and even values can be pointed out, but to do so is not necessarily useful. For example, Kohn (1969) found that social classes valued obedience differentially. This is an important study, but it would be a mistake to assume from this that clashes over obedience will occur in any one situation, such as a school classroom. Both lower and middle class children could be within the teacher's tolerance limits for disobedience and still differ enough to be statistically significant from each other in obedience. Although Kohn was particularly concerned with differences between the classes, he did explicitly point out that the differences he found were quite small and perhaps unimportant in many situations.

This ignoring of the degree of importance of a difference can be compounded methodologically several ways. First, some studies use an exceptionally large number of subjects. Theoretically, the probability of any two infinitely large groups being exactly the same on any scale is infinitely small. So the investigator needs only to take a large enough sample and he is bound to find some differences between the groups. For example, practically any two groups of people differ in height even if it is only in hundredths of an inch. With a large enough sample, such differences would be statistically significant but it is of trivial importance in the normal affairs of man.

A second methodological problem compounding evaluations of the importance of a difference lies in whether or not ipsative scales are used. An ipsative device, such as the rank order scale used in the present study, magnifies differences since it forces the respondent to choose between things which might be of equal or almost equal value to him. It can be easily misused as is done

whenever the differences between the ranks at the high and low ends of the scale are assumed to be equally distant as compared to those in the middle of the scale. Given the widespread prevalence of the normal curve in human affairs, this appears extremely unlikely. Therefore the differences would again be distorted.

Most of the studies have been carried out on readily available scales. These scales have seldom been developed from or analyzed on the groups being studied. The investigators assumed that the scales were functioning the same in all groups. Given the number of items which we found to be inadequate for one subgroup, this assumption appears questionable.

The importance of having a scale where each subgroup's reliabilities vary only randomly from each other can be underscored by following a hypothetical example. Assume that subgroup A has a normal mean and reliability for a values scale. In the present case, this reliability would be between .7 and .9. Since most people recognize the importance of the commonly accepted values, the mean is considerably above the midpoint of the scale.

Now consider group B. Assume that they have exactly the same values as do the members of group A. However, the language of the value scale confuses them because it is not a part of their experience; each child in this group gives up and answers randomly. Group B's reliability would be zero. Group B's mean's would then be the mean for random responding, which is at the mid-point of most value scales. Since this is different from group A's mean, a significant test may find the two groups' values to be significantly different even though we know they are not. The general principle is that a group with a lower reliability will have a mean different from a group of higher reliability whenever the latter's mean departs from the mean achieved by "guessing" or answering randomly.

The present section of this chapter reports the significant differences found on our value scales between the subgroups in Studies II and III. These are analyses of the non-ipsative scales which functioned adequately in each subgroup, but they do use large samples. The last section in the chapter discusses the meaningfulness of the significant differences.

Group Differences in Values: Study II

Some analyses of ethnic differences were possible from Study II with an abbreviated form of the Semantic Differential scale. Only four value definitions were used at that time; these were: Academic Achievement, Affiliation, Good Manners, and Kindness.

Subjects were randomly dropped so that an equal number of subjects per cell(14) could be entered into an SES by sex by race

by grade by semantic differential scales analysis of variance. The evaluation ratings of each concept were used. While there was no significant interaction in this analysis, there was a significant SES difference with middle class pupils scoring higher and a significant sex difference with females scoring higher. There was no significant race difference although blacks scored slightly higher than did whites. The difference between fifth and sixth graders was not significant.

The Value Socialization scale was also analyzed in Study II. Since the data in Study II were those used to select the items for the socialization scale, these analyses involved some capitalization upon chance. Fortunately, that capitalization occurs in terms of increased reliability for each of the subsamples. By the procedure used to select the items, it was not likely that there was any capitalization upon chance in the item means, but grade could not be included as a factor.

A three way analysis of variance was computed on the Value Socialization Scale. In addition to the main effects all being significant and in the expected direction (i.e., females, whites, and middle class children being higher), there was also a significant SES by sex by race interaction. A Newman-Keuls (Winer, 1962) test on the total sample showed the male and female white middle class pupils to be grouped at the high end of the scale whereas the white male lower class and female black lower class students were grouped at the opposite end of the scale.

Since the socialization scale showed some skewness, the results of the analysis of variance could be questioned. To determine whether or not the results would be significant by procedures not affected by the skewness, a median test was computed and run. The chi square was statistically significant ($p<.001$) and the results were quite similar to the analyses of variance. The middle class white females and males were grouped at the upper end along with the lower class white females whereas the lower class blacks grouped together with the lower class white males at the bottom of the scale.

Group Differences in Study III

For Study III, a series of analyses of variance was computed. The designs varied for different analyses, ranging from three factor to five factor designs. Scores for the eight values in the Semantic Differential Scale were treated as repeated measures. The number of subjects varied for different analyses and ranged from 96 to 397. Due to the skewness of the socialization scores that have been noted in the analysis of Study I, the F-max procedure (Winer, 1962) for checking the homogeneity of variances was computed on the Value Socialization Scale's primary analysis. It

suggested the analysis of variance could be legitimately run on this data since that statistic was not significant. Analyses of the order effect did suggest that some scores tended to be lower if the scale was given towards the end of the testing period rather than at the first. The order effect did not interact with any other dimensions for the socialization scale. All analyses reported here counterbalance for this order effect.

Order did interact with sex, race, and SES in the rank order scales. This produced some complex results with the conclusion to be drawn being based on whether the rank ordering took place early or late in the testing. On the basis of the confusing effects of order of administration on this scale, its preliminary nature, and its unknown reliability, it was decided not to report any of the analyses on subgroup differences.

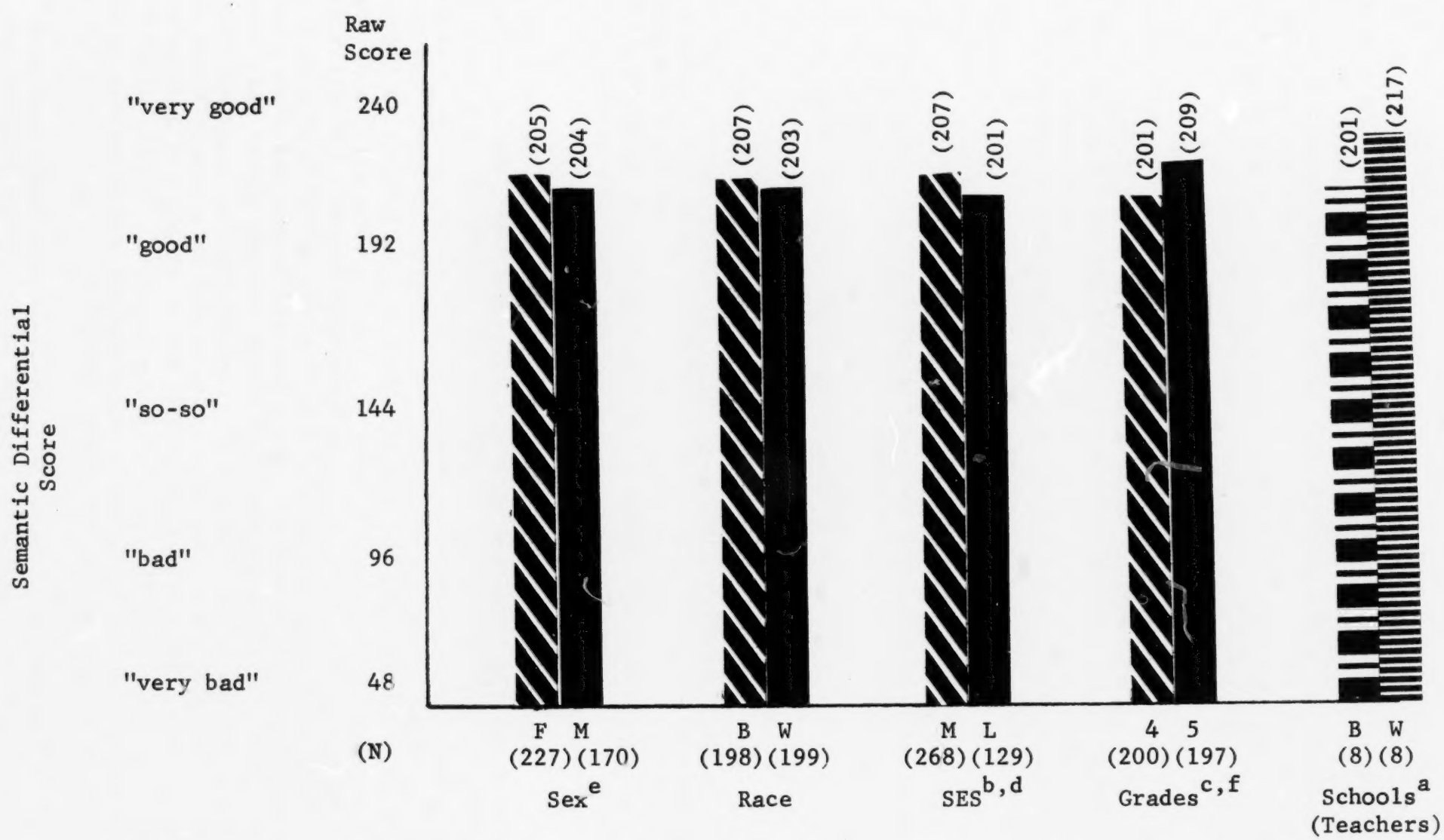
Results for the Value Socialization and Semantic Differential Total Value scales are easy to report since no interactions occurred. The effects of sex, SES, race and grade level did not interact with each other on either scale.

Neither did any of these dimensions interact with individual semantic differential value scales themselves. The lack of an interaction between the individual semantic differential value scales and any of the other dimensions means that each and every group preferred the same values relative to the other values. Physical Appearance and Development was rated the highest and Obedience the lowest (this was not the same order as that found with the Rank Order approach.) Any change in the rank order of these values when going to various sex, SES, race or grade subgroups can be attributed to random variations.

But the lack of any significant interaction between the scales and the subgroups does not mean that there was no differences whatsoever on the semantic differential scale. It only means that all differences found were consistent across all of the subscales for the various values. Significant mean effects did occur and these are given along with the non-significant main effects in Figure 5.1. As noted in the figure, neither the sex nor race effects on the Semantic Differential Total Value Scores were significant. The class differences were significant with middle class children scoring higher than lower class children. The grade difference was also significant with those children in the fifth grade scoring higher than those in the fourth grade. To aid in interpreting the meaningfulness of these data, their teachers' mean scores were also plotted on this scale. Those teachers in the white schools had average scores as high as any of the groups of the children. Those teachers in the black schools had a score which was not significantly different from the lowest of the childrens' groups.

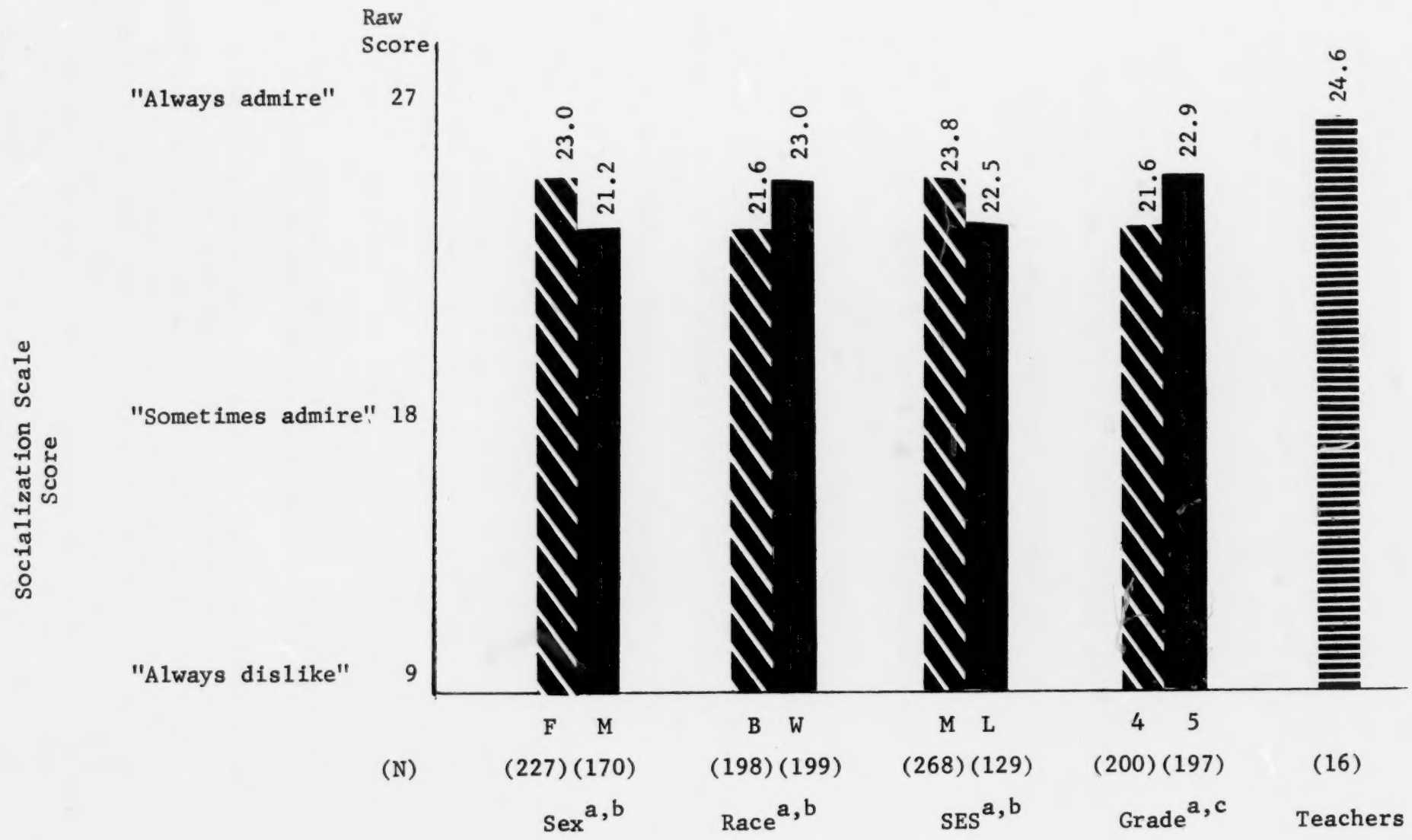
The main effects of SES, sex, race and grade were more striking on the Value Socialization scale. These results are presented in Figure 5.2. Again, the teachers are also included in this graph for comparison purposes but since the teachers did

Figure 5.1. Semantic Differential Total Value Means for Sex, Race, SES, and Grade Level Groups



[a]Difference statistically significant at the .06 level
[b]Difference statistically significant ($p < .05$)
[c]Difference statistically significant ($p < .01$)
[d]Replicated Study II's results.
[e]Significant in Study II.
[f]Not same test as in Study II.

Figure 5.2. Value Socialization Means for SES, Sex, Race, and Grade Level Groups



[a]Difference statistically significant ($p < .01$)

[b]Replicated Study II's results.

[c]Not tested in Study II.

not significantly differ among themselves, only their group mean is shown. The differences were all as common folklore predicted: females were higher than males, whites were higher than blacks, middle class were higher than lower class, and fifth grade pupils scored higher than fourth graders. Since the interactions were non-significant, this suggests that the male, black, lower class, fourth grade child was the lowest scoring child in this study.

The most consistent finding in these analyses was that the lower class child scores lower on any of the value scales than does the middle class child. It was also found that fourth graders score lower than fifth graders on these value measures. differences in sex and race were exhibited on some scales with the significant differences being in the expected direction. On other value measures, however, the significant results did not occur and there was occasionally a non-significant tendency for the groups to revise their order of scoring.

The Importance of Differences in Group Values for the School Setting

Most analyses of ethnic differences stop with the concluding paragraph given above. They simply note where the significant differences lie. This common procedure has already been criticized since it does not take into account the magnitude of the differences. The differences may be statistically significant but still negligible from any practical viewpoint, or they may be just barely significant but still of extreme importance due to their probable impact.

One approach to understanding the importance of the differences is to examine Figures 5.1 and 5.2 more closely in terms of the meaning of the ratings.

Looking at the means relative to the scale on which the children were rating the items provides a feeling for the nature of the responses. From the semantic differential, it can be seen that all children consider the values included in that scale as being one-fourth of the way from "good" to "very good", the highest point on the scale. No subgroup of children considers, on the average, the values set less than good. Each and every group of children is personally attracted to these types of behaviors.

And do these children admire the socialized, conforming child? The results of Figure 5.2 indicated that all of the children admire the child who shows the socialized behaviors included in the Value Socialization Scale. Each and every group mean is halfway between the category "Sometimes Admire" and that of "Always Admire", which means that the children may occasionally feel that the behavior described is inappropriate but such an occasion does not occur, in their eyes, very often. Therefore a subjective interpretation of the scores on both of these value scales indicates that all of the children accept the same common

core of values, a set of values which the traditional school setting is usually seen to hold.

This does not mean that there are not occasional individuals who do score much lower on each of these scales. This is certainly so but such individuals are almost as likely to occur in one group as in another group.

An objective approach to the importance of the difference between groups would be to ask how these groups differ on the behaviors predictable from values. By examining the later chapters, it can be seen that there are several places where values do predict behavior. By utilizing the group means reported in this chapter with the results of those later chapters, differences in behavior which are a result of value differences between the groups can be predicted.

Later chapters suggest that values have some impact on grades, the number of value conflicts in which the child is involved, and whether the teacher sees the child as a behavior or personality problem. From the predictive equations, estimates can be made as to how great a value difference must be to affect these criteria. But since both the criteria and the predictors are continuous, no single point can be an absolute division between different "types" of children. This means that no absolute decision can be made as to when the value situation is sufficiently negative to warrant drawing a line there. Any division established as the critical area could be legitimately challenged from another perspective.

Divisions are always for some purpose and involve areas of tolerance. Those areas will vary both with the situation and with the purpose. Therefore the estimation of when a value difference between groups is sufficiently important to warrant notice will legitimately vary as a function of one's perspective.

In our case, we take the traditional classroom which the project studied as the situation and ask what makes an important difference there. "Important" is defined somewhat arbitrarily. We started with the notion that it should be a difference which is reliable and interpretable. To be a reliable difference, the teacher's rating should be greater than any difference which might arise solely from rating the child on a parallel form of the criteria. The internal consistency reliability (coefficient alpha) estimates what that correlation would be. This coefficient was available for the behavior and personality problem checklists and suggested that an observed shift of three points would be likely (p of .05) to be an actual shift (i.e., more than twice the standard error of measurement, Nunnally, 1967). A score of three meant that the teacher found three more of the fifteen

adjectives on the behavior or personality problem list to be descriptive of the child. This also appears to be a "just interpretable difference."

Reliability data was available for neither the grades nor the value conflicts. Therefore the definition of an important difference relied soley on interpretability. We saw half a grade point average to be interpretable since this amount of change would move a child from, for example, a solid C to a weak B. Two value conflicts were assumed to be the minimum interpretable differences in that category.

The results of the later chapters were examined for differences in values which could predict a difference of half a grade point average, involvement in two value conflicts, three more checks as a personality problem, or three more checks as a behavior problem. The results naturally varied as a function of the value predictor and the criterion. (When the prediction differed significantly across groups, the median number of points on a value scale necessary to produce a significant and important difference was taken as the estimate.)

For the Semantic Differential Total Value Score, the teachers being forty-three points above a child would predict an interpretable difference in grade point average. A difference of sixty points would lead to a different prediction for the personality problem checklist.

An examination of Figure 5.1 shows the largest mean difference between teacher and the average child of a group to be only seven points. This amount of difference would not lead to a noticably different prediction for any of the criteria as a function of mean differences in group values.

For the Value Socialization Scale, eleven points difference in the child's score or in the difference between he and his peers predicted a difference in GPA. Being twelve points different from peers would predict a difference in value conflicts. Less of a difference from the teacher was needed to predict the criteria. If the child were five or six points below the teacher, then the child's grades, personality problem rating and value changes would be expected to be adversely affected; if he were four or so points above the teacher, then he would also be more likely to be rated a personality problem by the teacher. The mean differences reported in Figure 5.2 shows the group differences to be less than these critical points.

The differences between the groups are all less than the critical predictive points on either value scale, and so the average child of one group would mix well with the average child

of any other group. No teacher would give a meaningfully different rating to the average child of one group as compared to the average child of another because of their value differences. While the percent of children with predictions noticably different because of values would vary some across the groups, the majority of all the children of all the groups are predicted from values to be in the same area on the criteria. This suggests that the observed group differences are statistically significant but lead to negligible differences in the classroom setting.

CHAPTER VI

SHIFTS IN VALUES AND VALUE DISCREPANCIES OVER A SEMESTER

Stability and Change in Values

Due to the difficulties in carrying out longitudinal studies, little research has been conducted per se on the stability of children's values. The research that does exist suggests that there is developmental progression in those values. Kohlberg (1969) has found children tend to become more mature in their moral reasoning across the years. Kagan and Moss (1962) also reported a similar finding. While Peck and Havighurst (1960) are often quoted as finding a stability in character development, this stability was only in the child's position in the group; there was a mean shift with all the children tending to become somewhat higher on the measures of character as they matured. It therefore seems apparent that children's values do develop across time. However, little is known about the process or about the role of socializing agents in that process.

As part of the present project's attempt was to understand the impact of value discrepancies upon value conflict in the pupil-teacher relationship, the change that a child's values underwent across the semester of study was analyzed. A particular interest involved whether differential changes might be associated with group memberships.

From the present transactional, social psychological perspective, changes in the values of the children were considered only part of the issue. A total view would ask how those changes occurred relative to the situation in which the child was involved. If, for example, the child were in a situation which encouraged growth, much change could be expected. If, on the other hand, the situation did not encourage growth, then little change would be expected. In what follows, the source of "encouraging growth" was considered to be the interpersonal situation in the classroom. For this reason, the analyses of time trends were interpersonally oriented ones. The question centered on whether or not the child's values shifted towards or away from those of the teacher and peers rather than simply if they shifted.

In practice, differences between analyses of value changes based on raw value scores and analyses of change using value discrepancies between the child and teacher may be small. This is likely to be true when the teachers are relatively stable across the semester. Analyzing the discrepancies formed by subtracting the child's score at the first and second testing from a single

teacher's score means that one simply analyzes all the scores in a given classroom subtracted from a constant--the teacher's score. Any changes indicating growth or decline in a particular value would occur regardless of what score the teacher had received. But the use of value discrepancies does provide information on the possible source of value shifts that are found. If, for example, some of the groups of children are actually above the teacher on values and do not change, then this is an interpretably different situation than where other children are below the teacher and do change. Such effects are more likely to occur if the teachers themselves vary in their values.

In the present study, a teacher did form a relatively constant reference point for discrepancy scores of her pupils. Their value scores had significantly less variability than the children's scores on both the Value Socialization Scale and the Semantic Differential Total Value Score.

While it is possible that teachers' values would shift over the semester due to their experiences both in and out of the classroom, there was little evidence in the present teachers' data to suggest that this happened during this semester. Table 6.1 summarizes analyses of mean differences across the semester and presents test-retest correlations between the teachers' scores.

Table 6.1 Teachers' Values

Scale	September		January		T (df=15)	Test-Retest Correlation
	Mean	S.D.	Mean	S.D.		
Semantic Differential Total Value Score	208.9	16.2	209.1	18.3	.06	.71**
Value Socialization Scale	24.8	1.8	24.6	1.9	.44	.59*

*$p < .05$

**$p < .01$

Not only are the mean differences insignificant, but the test-retest correlations are significant even with this small sample of 16. On the basis of these analyses, each teacher's scores were averaged to form a basis for computing her children's discrepancy scores.

Note that there is also a theoretical reason to average the teacher's scores regardless of whether or not significant diff-

erences were found between testing sessions. The teacher's impact during the semester would be a function of the values that the teacher held during that semester. If her values did change, the average between the early and late testing sessions would still be the best indication of her average impact across the semester. While such a measure is, from our theoretical perspective, somewhat gross, it is one which was feasible for this project.

Changes in Pupil-Teacher Value Discrepancies

The Analyses

Analysis of variance with repeated measures for testing the changes over time was the major method of data analysis. Specifically, two major five factor designs with appropriate follow-up analyses were used. The first of these included race, sex, and grade as between groups factors and time and type of discrepancy (from peer and from teacher) as within groups factors. (A third within groups factor, individual values, was included for those analyses focusing on the specific values of the Semantic Differential). The second of those two major designs included the first, but also utilized teacher's estimate of SES as a between groups factor. This analysis, constrained by a small number of white males who were rated lower SES, was based on only one-third of the sample. Race, sex, and grade effects were based on the first set of analyses while SES effects and interactions with SES were based on the second. All subjects possible were used in each follow-up analysis to increase precision.

Main effects of sex, SES, race, and grade are not discussed here since they basically replicate the findings in Chapter V. The present discussion is principally concerned with time trends and with the time by groups interactions.

Time Trends

In the analyses of both the Value Socialization Scale and the Semantic Differential Total Value Score, significant changes were found in teacher-pupil discrepancy scores across the semester. Figure 6.1 presents the mean Value Socialization discrepancies as a function of time for the fourth and fifth grades. Figure 6.2 presents the same data but upon the Semantic Differential Total Value Score. On the former scale the change across semester was found in both grades but on the latter score there was a significant interaction between grades and time of testing during the year. Only the fourth grade showed a significant change. Note, however, that at the beginning of the semester the fifth graders' Total Value Score was higher than their teachers'. In contrast, the fifth graders' Value Socialization mean was still below the teacher mean in September and continued to shift. This

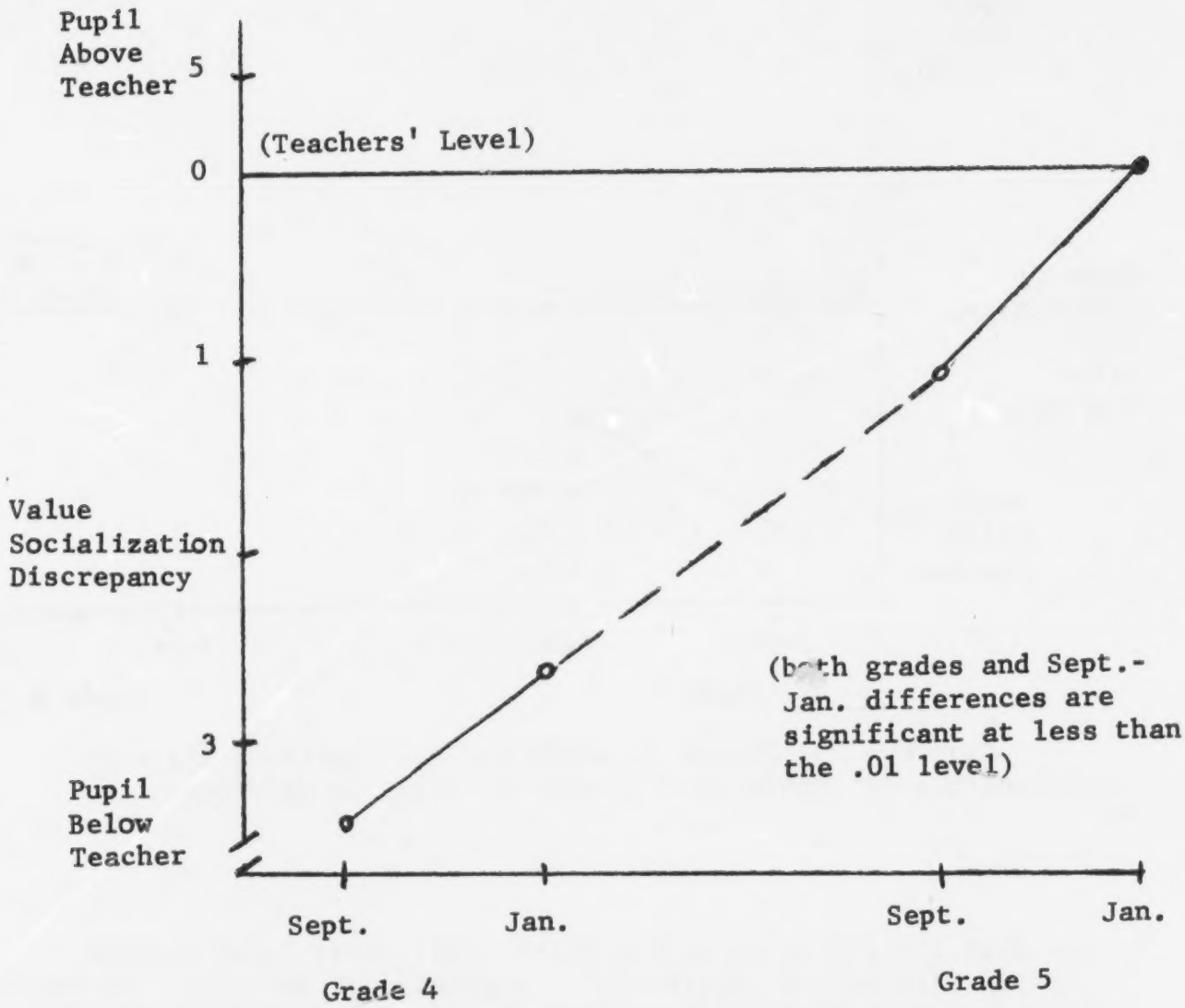


Figure 6.1. Change in teacher-child discrepancies over time and grade: Value Socialization Scale

suggests that a partial cause of shift may be the discrepancy between the pupil and teacher. When the children reached the teacher's level on a given instrument, they no longer increased.

Note that Figures 6.1 and 6.2 were laid out so that the eight month time period between January and September is indicated. The January value discrepancy scores for the fourth graders are connected to the September value discrepancies for the fifth graders. While this is not a longitudinal study and the subjects tested in the middle of the fourth year were not the same as those

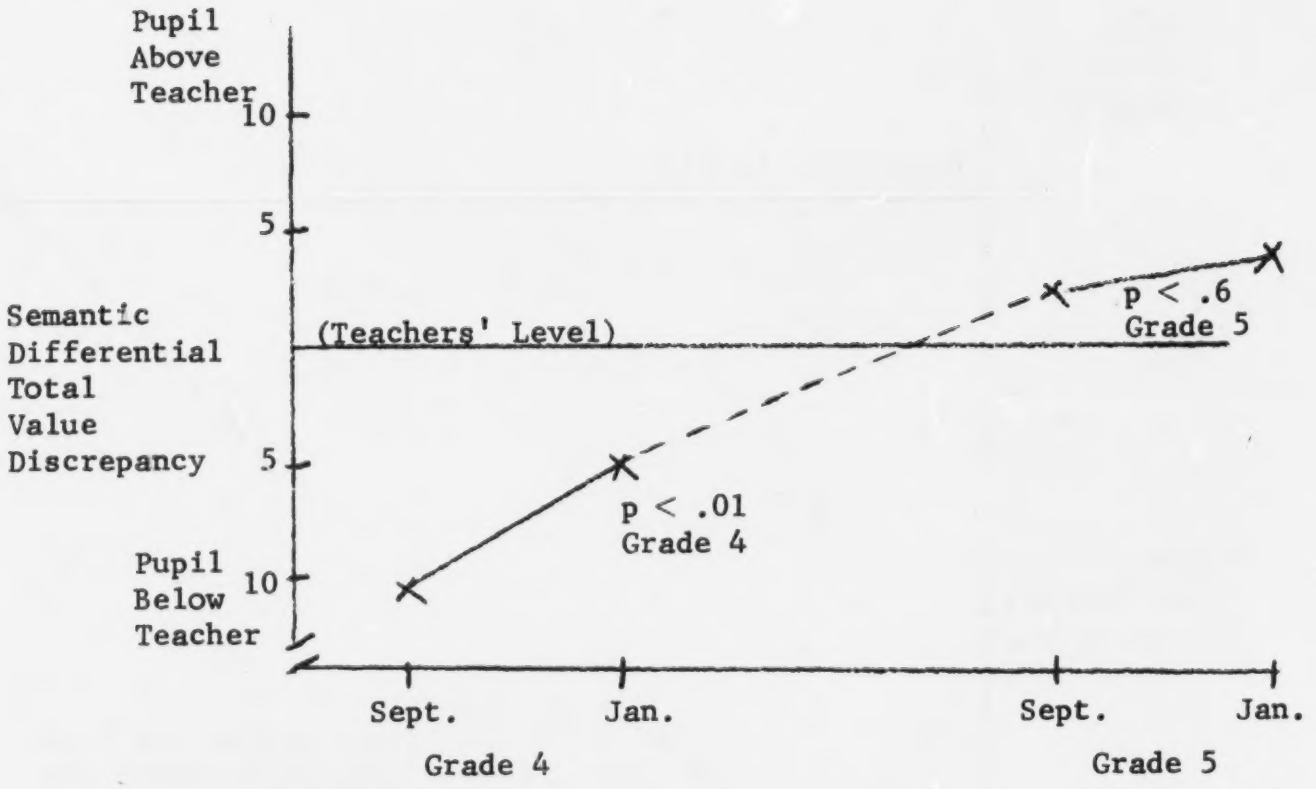


Figure 6.2. Change in teacher-child discrepancies over time and grade: Semantic Differential Total Value Score

tested at the beginning of the fifth year, these lines suggest that one alternative hypothesis is probably not accurate: children did not score higher in January simply because they were retested on the same scales. This alternative explanation provides no rationale for the fifth graders scoring above the fourth graders. Instead, the general developmental explanation seems to be more reasonable. As pointed out above, this is consistent with the general results of other developmental studies in the field.

Group Interactions with Time

The analyses of the Value Socialization Scale found no significant interaction between any of the groups and changes over time. Regardless of the sex, SES, race, or grade of the child, the results followed the graph given in Figure 6.1. There were main effect differences but these differences followed the pattern already discussed in Chapter V.

The analyses of the Semantic Differential Total Value Scores did show two significant interactions between membership in one of the groups and changes over time. Figure 6.3

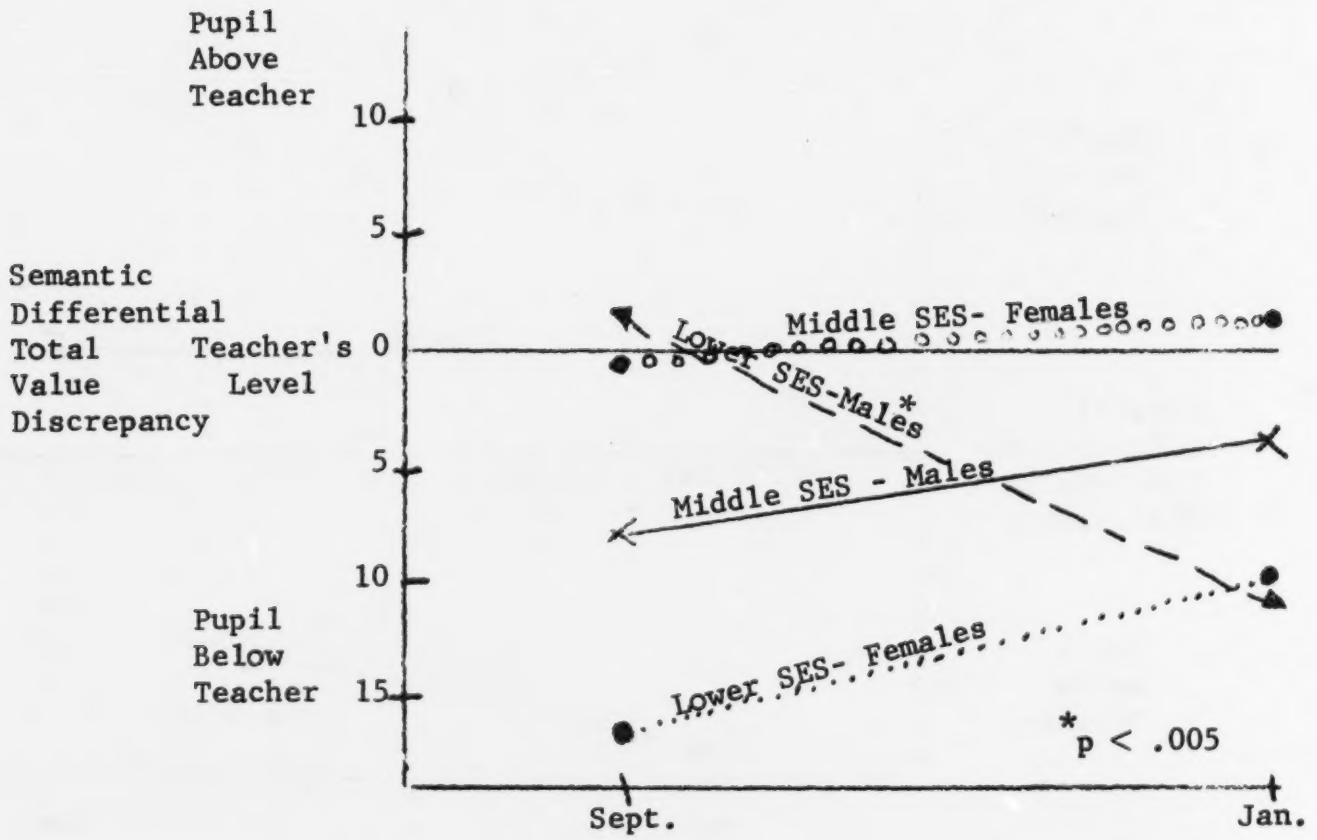


Figure 6.3. Sex and SES differences in change over time: Teacher-pupil discrepancies on the Semantic Differential Total Value Score.

presents aplot of the change over time in the Total Value Scores for each of the four groups formed by crossing SES and sex. Three of the four groups in that figure follow the general conclusions, but the fourth group, the lower class males, followed their own trend. They tended to increase in discrepancy from the teacher as the semester progressed.

Without further data, the exact cause of this movement is difficult to know. It is not purely a function of differences in teachers' level since lower class males were not taught by teachers who did not also teach lower class females. Perhaps the teachers in the schools with the greater percent of lower class students responded differently to males than to females. If this was the case and if such differential responding continued over a several year period, it would be no surprise if adult lower class males would not fit into the middle class culture which teachers generally

represent. They would not have been trained to fit.

Race also interacted with time in the Semantic Differential Total Value Score analyses. Figure 6.4 shows that interaction.

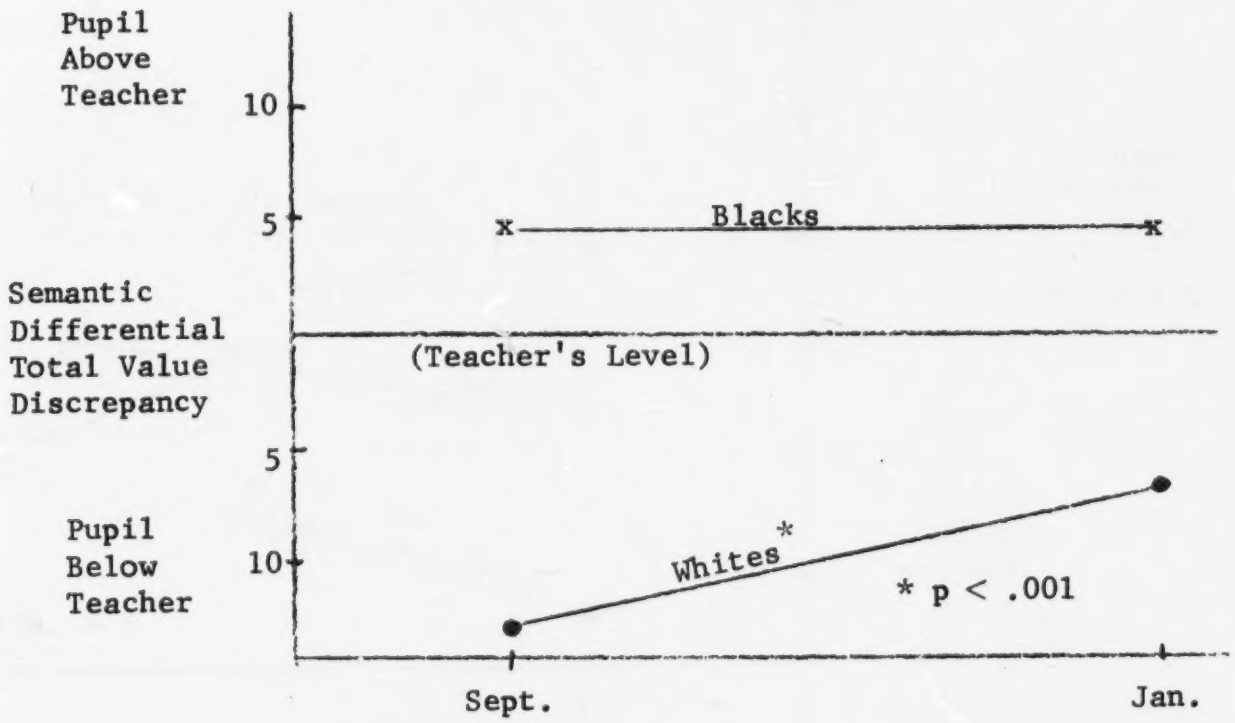


Figure 6.4. Racial differences in change over time: Teacher-pupil discrepancies on Semantic Differential Total Value Score.

Whites reduced their value discrepancies from the teacher as the semester went on, while the blacks did not change significantly on value discrepancy.

Figure 6.4 is, it must be remembered, a plot of discrepancies from the teacher. That score is a function of the teachers' value scores as well as those of the pupils. In this particular case, the graph suggests that blacks are higher in their values than the whites and this is not so. It only appears so because the teachers in the black schools were, as noted in Chapter V, lower on the Total Value Score than the teachers in the white schools.

Figure 6.5 plots the same effect but as a function of the pupil's absolute scores instead of as discrepancies from the teachers.

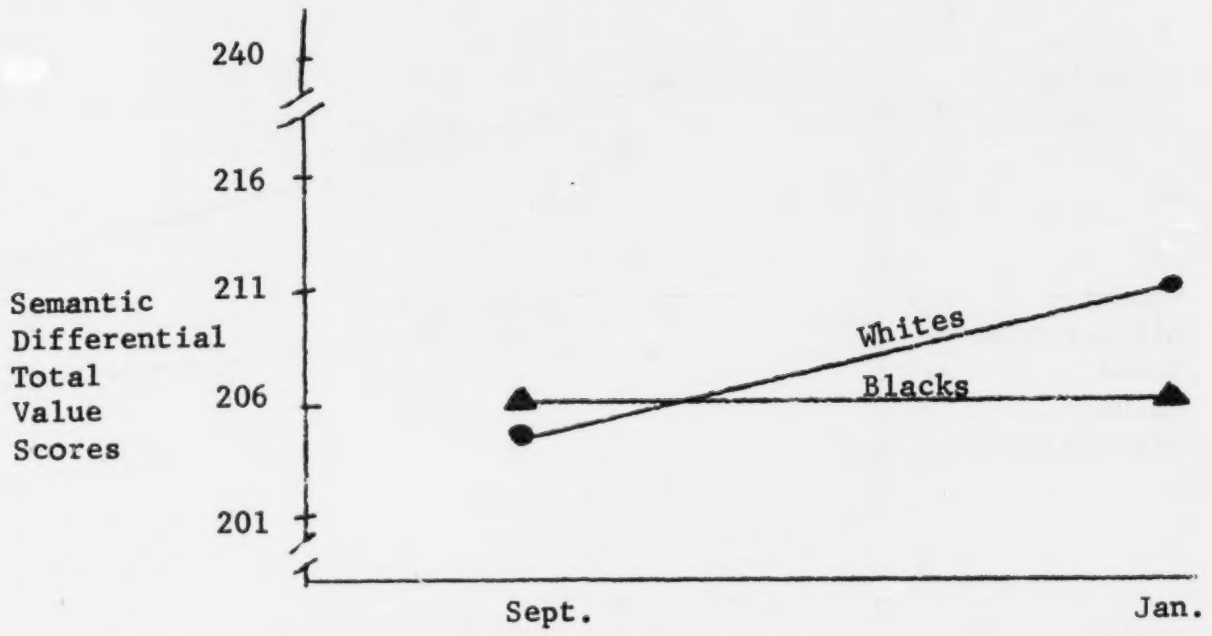


Figure 6.5. Racial differences in change over time: Pupil's Semantic Differential Total Value Scores

Figure 6.5 shows that while the whites and blacks started at about the same level, the whites increased their scores across the semester while the blacks remained the same.

These two figures jointly suggest that the teacher is having an impact upon the children's values. When the children are lower than the teacher, as was the case in the white schools, then they shift up towards the teacher. When the children were already as high as the teacher, then they remained relatively unchanged.

It is possible that the results in Figures 6.4 and 6.5 were a function of differential reactions across race rather than of the relationship between the pupils' initial positions and the teachers' initial positions. The plausibility of this interpretation, however, was reduced when additional analyses were run. All children whose initial scores were above their teacher's score were divided into blacks and whites. A race by time analysis indicated that there was no significant interaction between race and time, i.e., children of both races who scored above their teacher in September functioned the same. These results are plotted in Figure 6.6. Note that this graph must be interpreted cautiously. The trend down in time for those who were initially above the teacher may only represent regression to the mean effects caused by the way in which the subjects were formed into groups for the analysis. The regression to the mean effects did not, of course, exist for

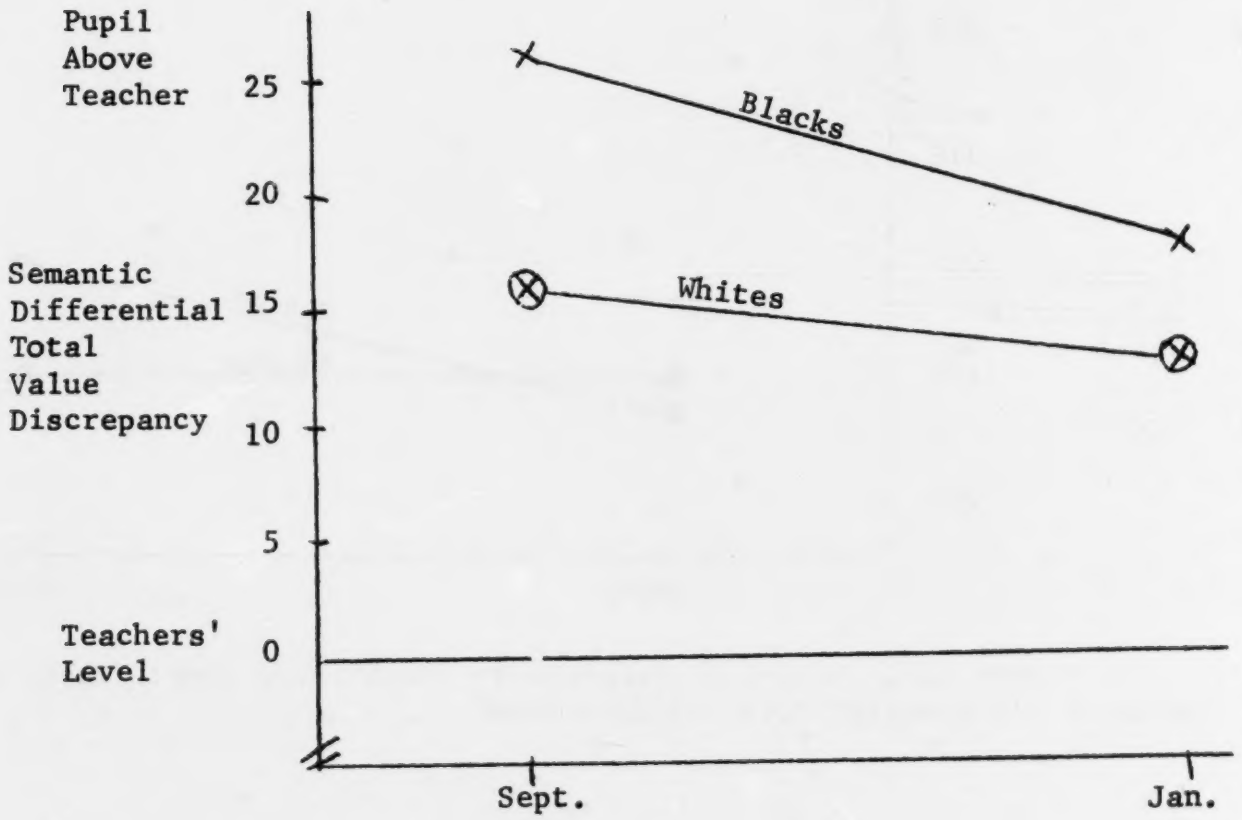


Figure 6.6. Racial differences in change over time: Subjects scoring above teacher only.

the another analyses presented earlier since subjects were not placed into groups on the basis of their value scores. In those analyses, there was no significant downward shift in values when children were above their teacher at the start of the semester.

Individual Values

The same analyses computed on the Semantic Differential Total Value Score were also calculated with the individual values. The general trend of these individual values was, as one might expect from the high correlations between them and the total score, consistent with the results found for the Semantic Differential Total Values Score. The exceptions that did exist were as follows:

(1) The average child in grade 4 was already slightly above his teacher on Good Manners and Physical Appearance and Development. Good manners did not increase further at grade 5 although there was some indication that Physical Appearance tended to increase even though the child was already above the teacher. Whereas most of the previous findings could be interpreted as possibly reflecting the impact of the teacher on the child's

values, the change in Physical Appearance suggests that other influences are likely to be the basis of the development of that particular value.

(2) The race by time interaction held true for all the values except that blacks and whites both increased on Academic Achievement. In almost every case where the September and/or fourth grade testing showed the children to be below the teacher on an individual value, the January and/or fifth grade testings showed those children to be much closer to the teacher. The children who were initially higher than the teachers usually remained higher.

Changes in Pupil-Peer Value Discrepancies Over the Semester

To obtain a measure of the discrepancy of a pupil from his peer group, the peer group was identified as those children in the classroom with whom the child had "run around" during the semester. The pupil's peers' value scores from the initial and the final testing session formed the basis from which that pupil's discrepancy scores were calculated.

In the analyses of variance where teacher and peer discrepancy measures were used, interactions occurred which suggested that these two discrepancy measures functioned differently. In follow-up analyses on the peer measures alone, it was apparent that only one significant pupil-peer discrepancy effect was found. On the Semantic Differential Total Value Score, pupils chose peers who were slightly above them in September and slightly below in January. This change in direction of discrepancy was significant. There were no interactions with SES, sex, race, or grade.

The overall level of pupil-peer discrepancy for both grades and both times on the Value Socialization Scale was significantly less than the pupil-teacher discrepancy. On the Semantic Differential Total Value Score, there was no main effect difference on these two discrepancy measures.

A Tantalizing Question: Sources of Change

The analyses noted above suggest a socializing agent effect on children's values. Generally, children tend to move up if their teachers score higher than they, but show little change if they are already as high as their teacher. (The lower class males were the one exception). This teacher effect was found when the children were basically at the same point of origin but had teachers with different scores (cf. Figure 6.5). Children with teachers higher than they increased in their values. Children with equal initial scores but without such teachers did not increase. It suggests

that a child who has a series of teachers who are high on the measures of values used in this project would score higher on those scales in the upper elementary years than would a child who had teachers below him during the same years. Our data also suggest that the conformity of most children to these sets of values had essentially occurred by the fifth grade.

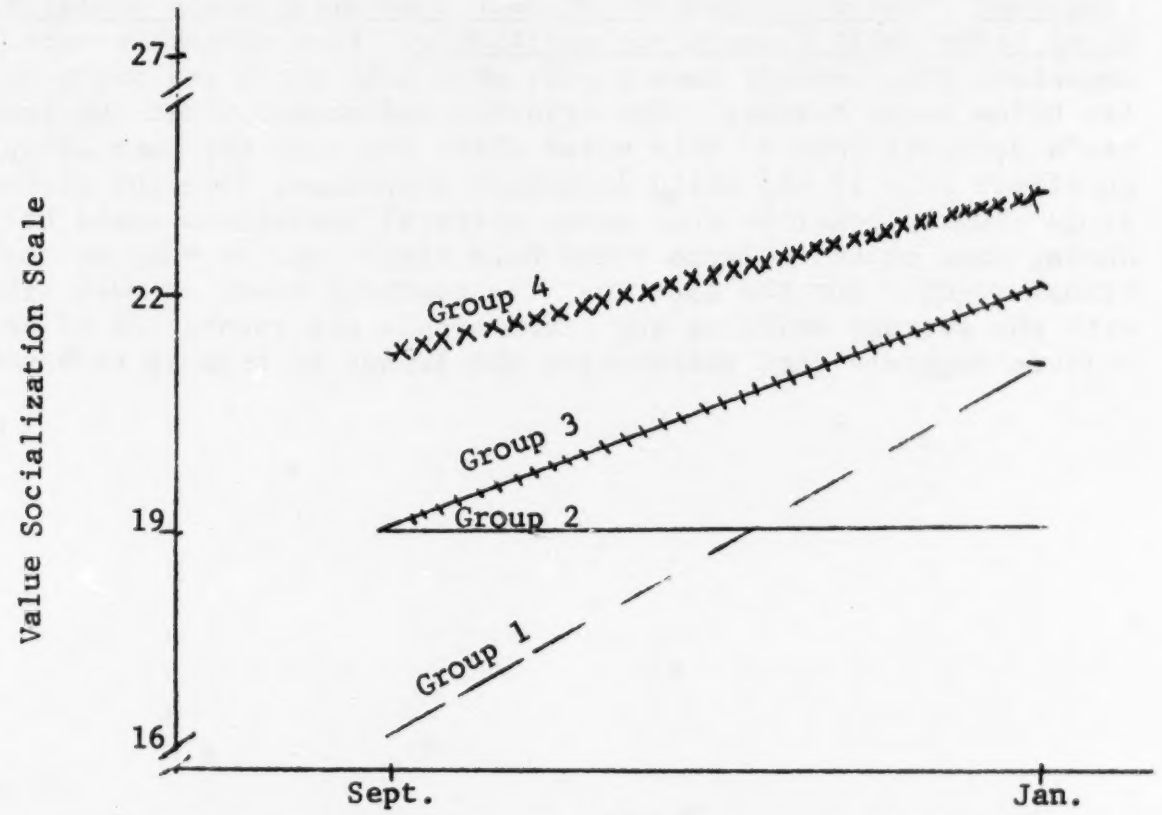
Conclusions regarding the peer group impact are somewhat more difficult to make. The peer group was only defined in one manner for these analyses, i.e., by the child's recollection of his membership group across the semester. A considerable amount of psychological theory (e.g., Sherif & Sherif, 1964) has suggested that the child's reference group may be more important than the child's actual membership group. No reference group analyses were made in this project.

To examine further whether or not the teacher's impact upon the children was possibly moderated by the peer group impact--or vice versa--a series of analyses of variance were run. Children were grouped according to initial Value Socialization discrepancy scores from the teacher (i.e., above the teacher, same as the teacher, moderately below the teacher, and far below the teacher) with similar distinctions being drawn on the basis of the initial peer group discrepancy scores. The correlation between teacher and peer discrepancy prevented the computation of a two factor analysis of variance, so a series of one way analyses of variance examined changes across time in these various groups, i.e., changes on the raw score over time for different peer and teacher discrepancy combinations.

Since the groupings were based upon the initial testing session, regression to the mean effects were expected. Interpretable finding could occur only when a group's change over time went against the regression to the mean predictions.

Such non-regression to the mean effects did occur in one important case. Figure 6.7 shows it contrasted with three other groups which did show possible regression to the mean. Group 1 were children with initial Value Socialization scores more than six raw points below the teacher and more than three raw points below their peers. Group 2 had initial scores at least six raw points below their teacher but less than three raw points below their peers. Group 3's scores were more than six points below their teachers but these children did not run around with anyone else in the class. Group 4 contained children with initial scores between two and six points below their teachers and less than three points below their peers.

The results were as one might expect from studies such as Roff and Sells (1968). Those children who were far from the teacher and had peer groups support at that position (Group 2)



Group 1: Score is 6+ points below T and 3+ points below P

Group 2: Score is 6+ points below T and same as peers

Group 3: Score is 6+ points below T and no peers named

Group 4: Score is between 2 and 6 points below T and less than 3 points below P

Figure 6.7. Change in value socialization over time as a function of teacher and peer discrepancy.

did not increase in their Value Socialization scores over the semester. Those children equally low but who either had peers closer to the teacher or had no peer group support within the classroom shifted towards the teacher.

The analyses presented in this chapter suggest that changes

in values do occur and that, for most children, the primary way of resolving value discrepancies between themselves and a socializing agent is to shift towards the socializer. Peer impact is more important than teacher impact only when both pupil and peers are far below their teacher. The evidence did suggest that the teacher has a definite role in this value shift but that the peer group has an effect only if the child is highly discrepant from the teacher. It is clearly possible that other cultural conditions could be producing some of the effects noted here since this is only an observational study. But the fact that the teachers spend as much time with the average child as any other single representative of adult culture suggests that attributing the impact to them is reasonable.

SO 101

CHAPTER VII

TEACHER PERCEPTIONS AND LABELING: AN APPROACH TO CONFLICT MANIFESTATIONS

Introduction

To evaluate the effect of value discrepancies in the school setting, it was necessary to have measures of conflict manifestations. Some measures were readily operationalized. For example, anxiety was defined phenomenologically by the child's feelings of discomfort. Since another's feelings can only be known by interpreting his reports of them, an anxiety questionnaire was given the children. Value changes and grades were equally operationalizable.

Other manifestations were more difficult to record. One procedure would be to have observers in each classroom for designated time intervals to report the value conflicts, other incidents, and what the children were like. Such a procedure would be, however, expensive. A less costly and less elegant procedure was adopted: the teacher reported her views and these were considered measures of conflict manifestations.

Teacher reports of important incidents and ratings of children are far from objective. Instead, they represent the teacher's perceptions or labels. Since these would form the critical dependent measures of the project, how the teacher developed and applied her labels was analyzed.

An analysis of the perceptual process has other implications as well. From the transactionist perspective, the perceptions a teacher develops are more than just evaluations. They also guide her in carrying out her role. A further understanding of these perceptions thus has implications for how the teacher might function relative to her students.

The Development of Teacher Perceptions: A Model

The semester under study was considered as a time during which the relationship between the teacher and the students developed. Such an acquaintanceship process has the following three phases:

Phase 1. Generalized Perceptions.

The teacher views the students as one does newcomers, on the basis of easily observable cues only, and uses global categories for judgment. This would be seen as stereotyping

from overt (superficial) cues. Here perceptual labels are for the purpose of structuring initial relationships and to provide assumptions about how to start behaving.

Phase 2. Perceptual Shifts.

Over time, the teacher's experience with pupils leads to the use of new categories and differentiation of global categories. The teacher would most likely be exploring how to view the students and would be attentive to pupil behavior to discern patterns of consistency. Here labels describe developing perceptual/labeling patterns.

Phase 3. Stabilized Perceptions.

With sufficient experience, a stabilized set of conceptual labels are developed. Once these are in place, the behavior of the teacher should settle down and become consistent. During this phase the labels provide the meaning framework and the assumptions on how to behave towards the child.

The perceptual process that the individual goes through is viewed analogically to a psychophysical one where perceptual inputs are compared against standards and are categorized as the same or different. The perception is regarded as having crossed into a new category when it is judged as different. When a category fit is achieved one experiences a concomitant positive or negative affective reaction. A negative reaction indicates a threshold of tolerance has been crossed. Operationally this is defined as behavior emitted by the perceiver to change or avoid the labeled behavior.

One measure in this study attempted to record the crossings of the threshold of tolerated behavior. This was a method used by the teacher to note the situation and her own reactions. The instrument was termed Important Incidents and was based on the Critical Incidents analysis of occupations developed by John Flanagan (1954). Important incidents were events that the teacher thought to be important or remarkable. She was asked to report the following on each incident (cf. Appendix E):

1) who was involved,
2) date and time of day,
3) what happened,
4) what she did, and
5) the reaction of the other pupils.

The negative incidents were then defined as conflict events, with "conflict" designating a negatively viewed discrepancy between expected and observed behavior. We decided to regard the negative labeling of value laden events as behavioral measures of value conflict. In this chapter, we are interested in the variables measuring teacher judgements in order to discern the meaning of label shifts and to see whether general expectancies change.

The Measurement of Teacher Perceptions

Teacher perceptions consisted of rating scales which measured different parts of the process and which required judgment on the part of the teacher. The measures and their hypothetical relationship to the process are summarized in Table 7.1 The measures are described below.

1. Self-Anchoring Scale. The Self-Anchoring Scale was developed by Kilpatrick and Cantril (1960) as a means for measuring the perceptions of the raters as compared with their own standards. The procedure called for the teacher to first define (in her own words) the "ideal student" and the "worst student." These served as the anchors for each end of the nine point scale. Second, the teacher was asked to rate all her students on this scale.

2. Value Ratings. The teacher judged each child's values by rating him on a nine point scale for each of the eight value categories carried in the project.

3. Problem Checklists. Peterson (1961) analyzed teacher's views of their children's problems by means of a checklist. Teachers' perceptions of deviate behavior, it appears, are concerned with two different aspects: behavior problems and personality problems. These two dimensions appeared in analysis of data from all grades. The resulting checklist has a set of adjectives for each area. The internal consistency reliabilities for the two scales ranged from .8 to .95 when calculated on data of Study II.

4. Important Incident Reports. This procedure was described above. It provided a non-rating report of actual events and was considered to be the only measure of overt value conflict. By the use of this measure, it was possible to have a clear distinction between value discrepancy and value conflict.

To determine whether or not an incident involved a value conflict, the raters read each incident and placed it into one of nine categories. These nine categories consisted of the eight values and an "other" category. Each incident was also rated as to whether it was positive, negative or neutral.

Table 7.1

Teacher Perception and Labeling of Behavior as an Acquaintanceship Process

Measures	General Expectancies	Phase I	Phase II		Phase III
		General Impression	Perception & Labeling	Perceptual Shifts	Stabilized Perception
Self-Anchoring Scale & Value Ratings	Anchor Descrip-tions	September Ratings		November Ratings	January Ratings
Problem Checklist		September Ratings		November Ratings	January Ratings
Important Incidents			Type & Frequency		

Value Conflicts were defined as negative incidents involving one or more of the values defined by the project. The percent of agreement for the two raters was as follows: Positive Incidents 66%, Negative Incidents 91%, and Neutral Incidents 81%. The percent of agreement for each of the categories was: Non-Value 74%, Academic Achievement 57%, Affiliation 18%, Diligence 43%, Good Manners 75%, Honesty 75%, Kindness 41%, Obedience 44%, and Physical Development 44%. The overall percent of agreement was 64%. This was not high and probably reflects the fact that most incidents involved more than one value category. Another reason for the unreliability probably arose from the fact that the categories were developed for a different purpose. To increase the reliability, a third rater resolved disagreements.

Each pupil's score was the number of times he had been involved in a particular type of incident. Note that under this scoring system a given incident could lead to several children being scored. This was felt to be appropriate since it gave slightly more weight to those incidents with more participants. No distinction was made between a person who was offended and the offender since, as in international politics, it would be difficult to ascertain with an acceptable degree of certainty.

Analyses of these measures were carried out by several techniques to determine the initial structure of teacher perception and how those perceptions of pupils changed over time.

The Structure of Perception: Factor Analysis of Teacher Ratings

It is generally conceded (e.g., Kleinmuntz, 1967) that most interpersonal ratings are global in nature. Little differentiation is made and, instead, people are simply judged as good or bad. To provide for intelligent use of the teacher perceptions, an examination of the extent to which they contained a global judgment dimension was made.

To study the extent of such a generalized perception effect, a sequence of factor analyses were applied to the Self-Anchoring Scale ratings, Behavior Problem ratings, Personality Problem ratings, and the eight value scores. First, communality estimates were computed for three factors since three to four factors maximized the amount of variance accounted for relative to the number of factors. The communalities were estimated by the iterative principal axis procedure (Harman, 1967). Second, the eleven variables were entered into a diagonal factor analysis (Nunnally, 1967) where the Self-Anchoring Scale defined the only factor to be extracted. This analysis tested the "generalized perception" hypothesis since the Self-Anchoring Scale specifically required a global/evaluative judgment. The resulting factor structure appears in the first column of Table 7.2. As can be seen from this table, a pronounced general factor did seem to exist. Eight of the eleven variables loaded highly (.70-.93) while ratings of Personality Problems, Behavior Problems, and Affiliation loaded moderately (.50-.52).

As a next step, two principal axis factors were extracted from the residuals of the diagonal analysis and rotated to the Varimax position (Harman, 1967). This tested for the existence of more specific perceptual/labeling constructs not accounted for by the Generalized-Perception factor. The resulting factors appear in columns 2 and 3 of Table 7.2 and are relatively small. Nevertheless, these structures were distinct in that the first residual factor was defined primarily by the Behavior Problem ratings (.61) while Personality Problems had the highest loading (.44) on the second residual factor. These replicated Peterson's (1961) dimensions.

Since the teacher's ratings of values appeared to provide no information beyond that given by the three factors, they were excluded from further analyses. Also note the joint loadings of the personality and behavior problem measures on this first factor; this suggests that the Self-Anchoring Scale may not be necessary if both problem scales are included in a particular analysis.

After extracting three factors as representative of the teacher's perceptual set (or structure), two important criteria--grades and value conflicts-- were correlated with them (cf. Table 7.2).

Table 7.2

Results of Factor Analysis of Teacher Ratings

Variables		Generalized Perception Factor Structure	Residual Factor Structures	
A.	In Factor Analysis			
	Personality Problem	-.50	-.15	.44
	Behavior Problem Rating	-.52	.61	.25
	Self-Anchoring Scale Rating	.80	--	--
Value Ratings	Academic Achievement	.93	.13	-.07
	Affiliation	.50	.34	.04
	Good Manners	.76	-.38	.26
	Physical Development	.72	.05	.31
	Obedience	.90	-.30	.14
	Kindness	.70	-.37	.06
	Honesty	.83	-.32	.08
	Diligence	.86	-.03	.00
B.	Correlated with Factor Scores			
	Grades	.62	-.49	.32
	Value Conflicts	-.64	-.24	.36

The moderate loadings of these variables on the generalized-expectancy factor suggest, first, that value conflicts are closely related to a global/evaluative perception of the pupil. Second, this global rating of best or worse student included judgments of pupil academic competence as found in grades.

Perceptions of Pupils Grouped by Social Class, Sex, and Race

The findings reported here and in the remaining sections of this chapter summarize the statistically significant results of a series of analyses of variance computed for Self-Anchoring Scale ratings and ratings of Behavior Problems, Personality Problems, and values.

Social Class Differences

The teacher's estimate of social class was used in these

analyses. This analysis therefore relates two sets of teacher perceptions. It is not, of course, known whether a significant relationship reflects an actual characteristic of that social class or whether the teacher places the child in that class category because he was seen as having the characteristic. It should be noted, however, that despite their subjective nature, teacher estimates were a reasonably accurate measure of social class. The findings were:

1) Self-Anchoring Scale. Middle-class pupils were rated higher than were lower-class pupils. (Figure 7.1).

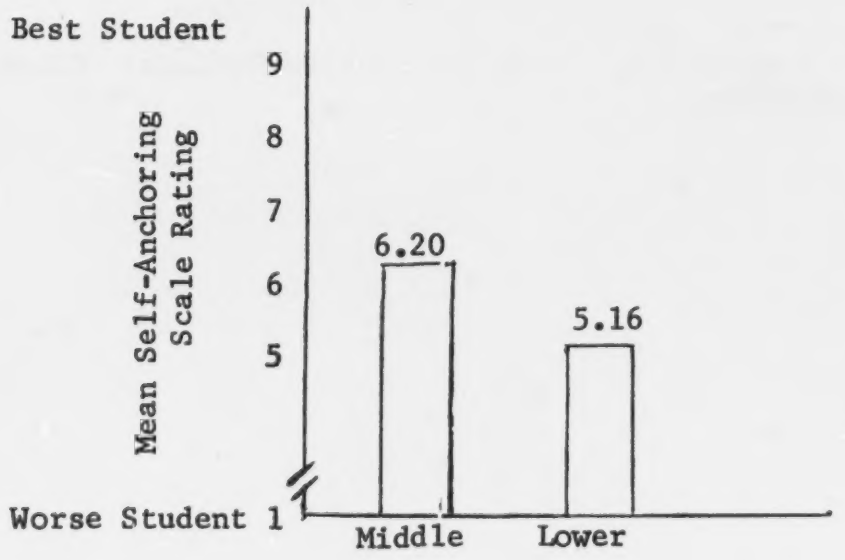


Figure 7.1. Self-Anchoring Scale Ratings: Class effect.

2) Behavior Problem Ratings. Generally, lower-class pupils received moderate ratings as behavior problems while middle-class blacks received the highest ratings and middle-class whites the lowest. (Figure 7.2).

3) Personality Problem Ratings. A rating pattern similar to that of the Behavior Problem ratings was evident. Middle - class blacks were rated highest, lower-class pupils were moderately high and middle-class whites were lowest. (Figure 7.3).

4) Important Incident Reports. Middle-class and lower-class pupils received, on the average, an identical number of Positive Incident reports; however, middle-class pupils received three times more Negative Incident reports than their lower-class

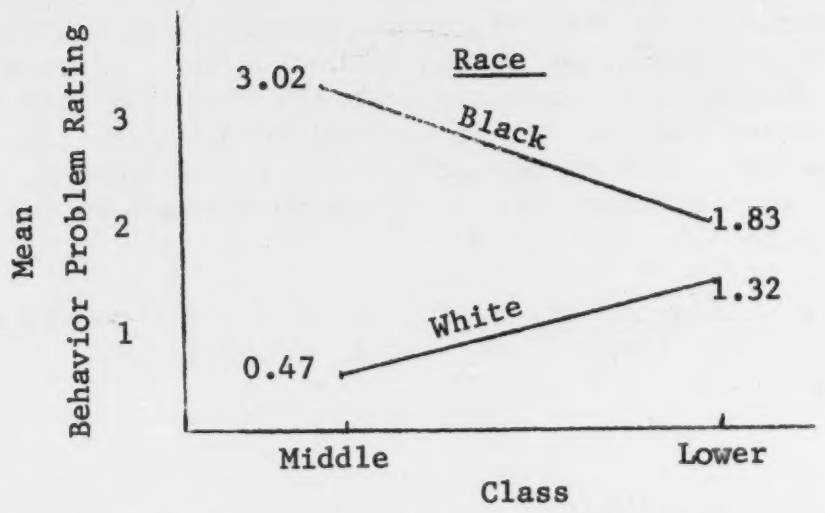


Figure 7.2. Behavior Problem Ratings: Class x race interaction.

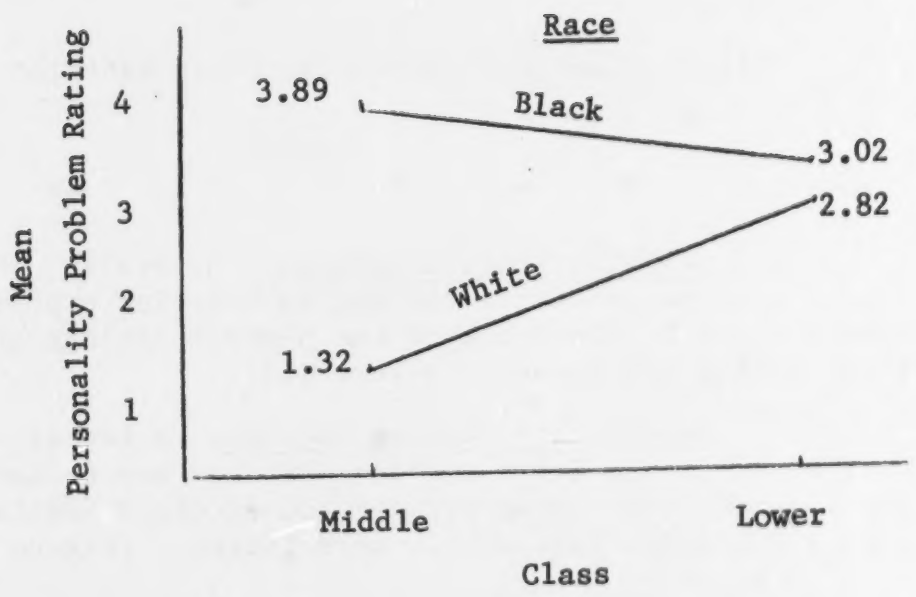


Figure 7.3. Personality Problem Ratings: Class x race interaction.

counterparts. (Figure 7.4). Note that this goes against most folklore. (For Value Conflicts alone, cf. Chapter 8).

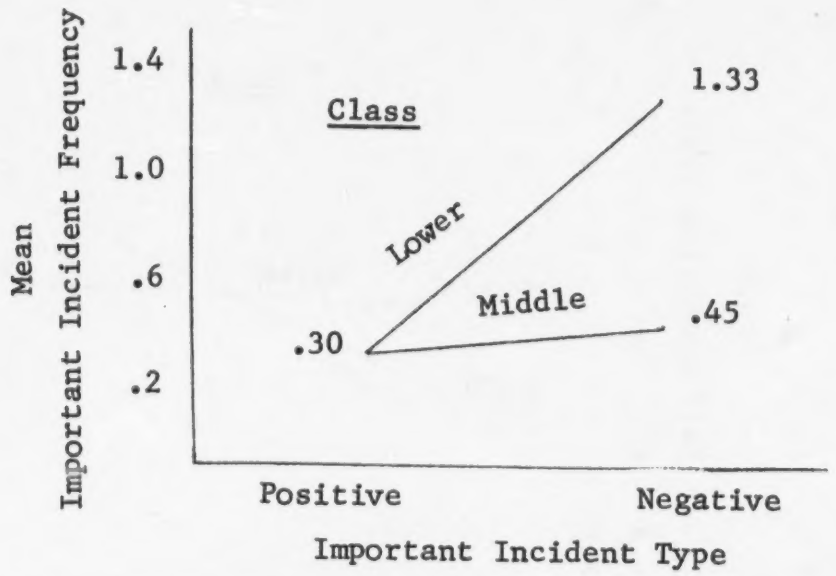


Figure 7.4. Important Incident Frequency: Class x type incident interaction.

Sex Differences

A more favorable rating of females than males was generally found.

1) Self-Anchoring Scale. White males and females received similar ratings; on the other hand, black females were rated higher than black males. (Figure 7.5).

2) Behavior Problem Ratings. Males received a rating more than three times as high as that for females. (Figure 7.6).

3) Personality Problem Ratings. Analyses on Personality Problem ratings revealed no sex differences.

4) Important Incident Reports. Males received slightly more positive incidents than females but a much higher number of negative incidents were reported for them. (Figure 7.7).

Race Differences

Results generally showed a more favorable rating of whites than blacks.

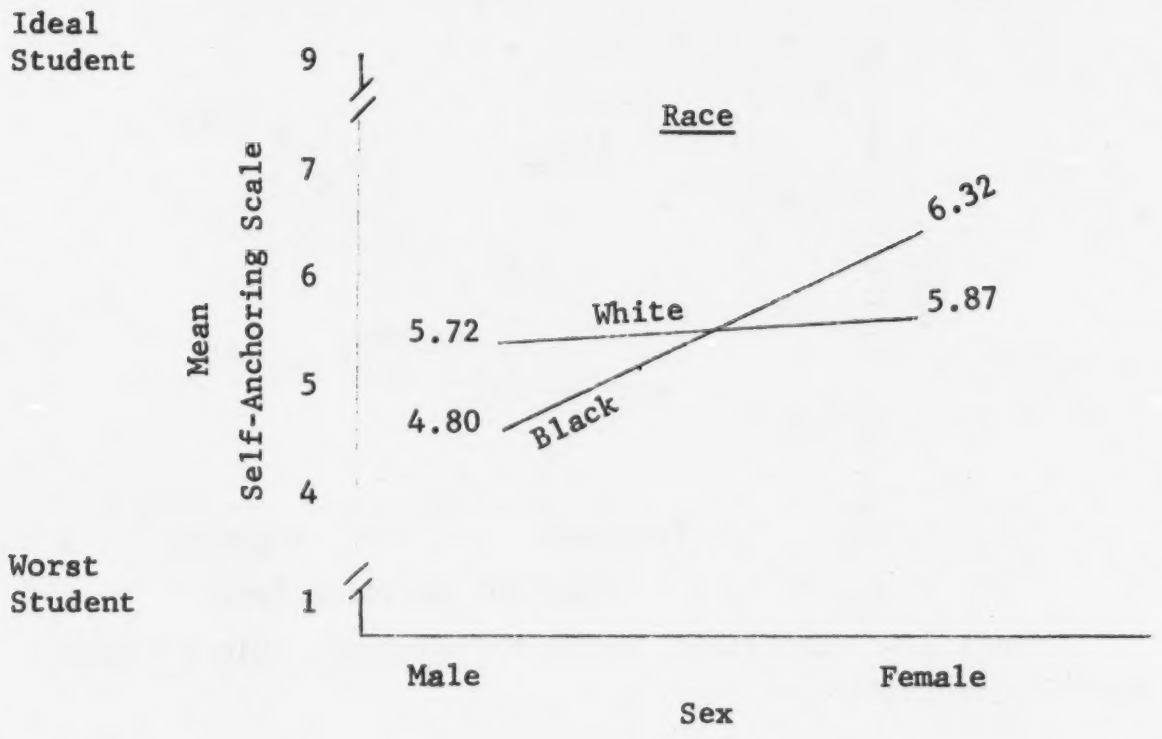


Figure 7.5. Self-Anchoring Scale Ratings: Race x sex interaction.

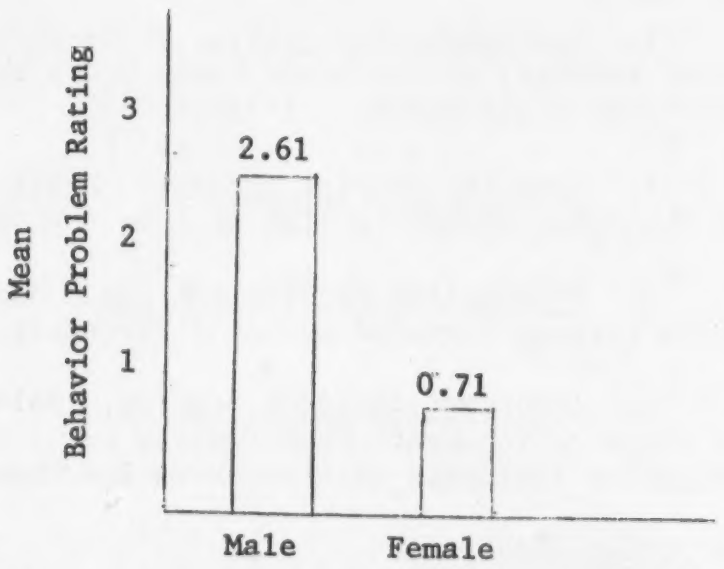


Figure 7.6. Behavior Problem Ratings: Sex effect.

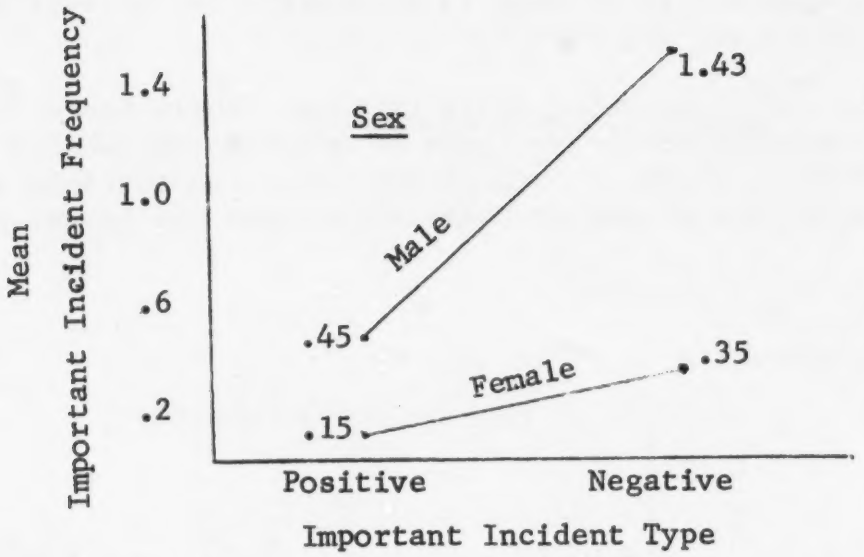


Figure 7.7. Important Incident Frequency: Sex x type incident interaction.

1) Self-Anchoring Scale. Whites and black females received moderately high ratings while black males are rated the lowest. (Figure 7.5).

2) Behavior Problem Ratings. In general, blacks were rated higher than whites; however, the difference seems to be a middle-class phenomenon. Middle-class blacks received a much higher rating than middle-class whites while lower-class blacks and whites were rated similarly. (Figure 7.2).

3) Personality Problem Ratings. Personality Problem ratings showed a pattern similar to that of behavior problem ratings. Lower-class blacks and whites were rated similarly with middle-class blacks rated much higher than middle-class whites. (Figure 7.3).

4) Important Incident Reports. No racial differences were found in analyses of incident frequency.

Perceptual Shifts: A Time Perspective

There were some changes in teacher perceptions across the semester as viewed through our three-point time-sampling procedure. These analyses sought to isolate time interactions with social class, sex and race and to examine relationships among the several measures of teachers' perceptions across time. These showed

interesting and yet puzzling results; tne pattern that emerged usually showed similar ratings in September and January with either a marked rise or fall in November. The effects on each scale appeared as follows:

1) Self-Anchoring Scale Ratings. While there was no overall race difference in terms of mean scores, whites were rated lower in November than at the other two testings with blacks showing a slight decrease throughout the semester. (Figure 7.8).

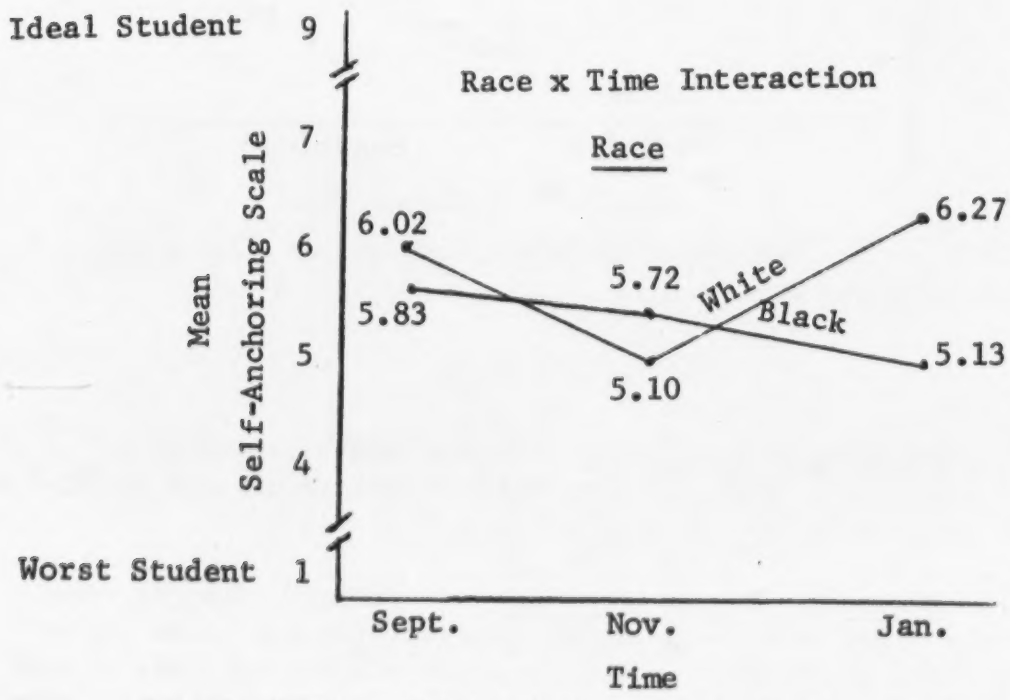


Figure 7.8. Self-Anchoring Scale Ratings: Race x time interaction.

Those pupils rated as both a Behavior Problem (6-15 group) and Personality Problem (7-15 group) showed a consistent decline in Self-Anchoring Scale ratings throughout the semester. (Table 7.3). Pupils rated as having neither a behavior problem nor a personality problem showed a much higher rating in November than at the first or last testing. In general, those pupils receiving no incident reports were rated slightly higher than those receiving all positive reports; pupils receiving all negative reports were rated the lowest. (Figure 7.9). The No Incident group and the All Negative Incident group showed a marked fall in November with the September and January ratings more similar.

Table 7.3

Self-Anchoring Scale Ratings:
Behavior Problem x Personality Problem x Time Interaction

TIME	Personality Problem Rating	Behavior Problem Rating: 0	1-5	6-15
September	0	6.57	6.14	4.57
	1-6	5.86	5.86	4.43
	7-15	6.28	5.71	5.14
		1	2	3
November	0	5.14	6.86	3.28
	1-6	7.00	4.43	4.43
	7-15	4.71	4.86	4.86
		1	2	3
January	0	7.86	6.43	3.00
	1-6	5.43	5.71	2.57
	7-15	5.14	5.14	1.00

2) Problem Ratings. Pupils receiving all Negative Incident reports were rated as more of a behavior problem and show a marked increase throughout the semester. (Figure 7.10). Pupils receiving all Positive Incident reports showed a slight peak in November while those receiving no incidents evidenced a slight decrease from September to January.

Teacher Labeling Style and Important Incidents

Negative Incident Reports have been defined as a criterion of value conflict since they directly reflect critical behavioral interactions in the classroom. The factor analytic data presented earlier illustrated that these value conflicts are closely related to a general global/evaluative factor which includes most of the perceptual dimensions used in the project. This section focuses

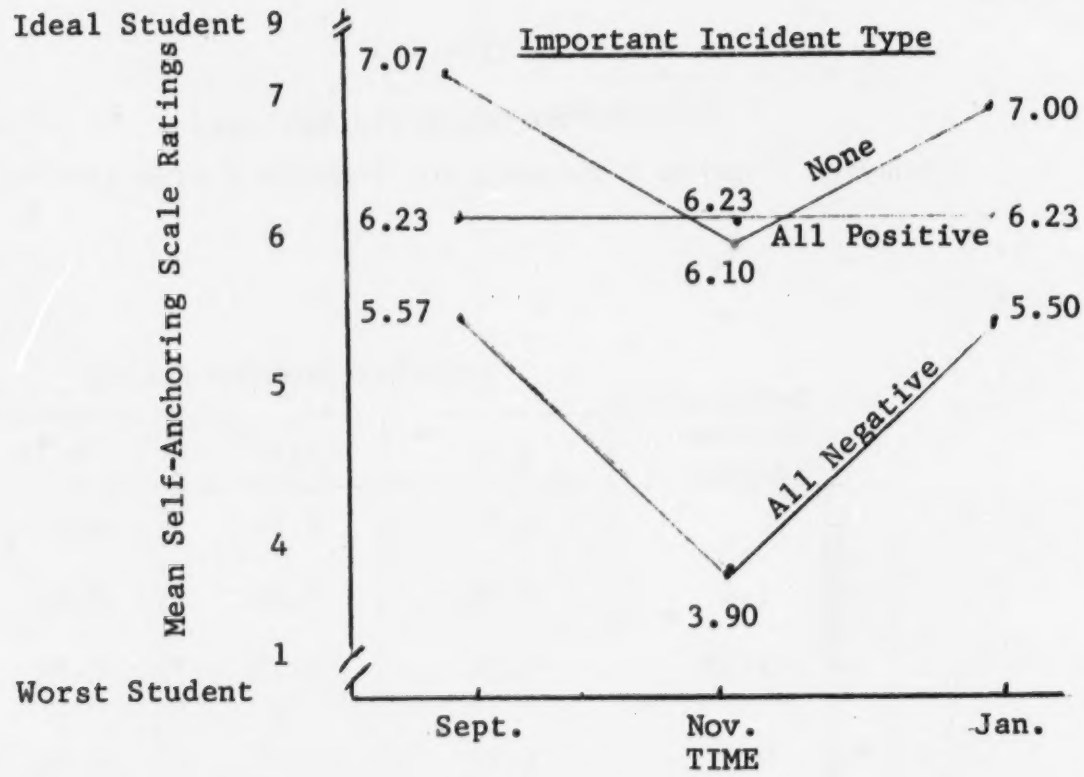


Figure 7.9. Self-Anchoring Scale Ratings: Important Incident type x time interaction.

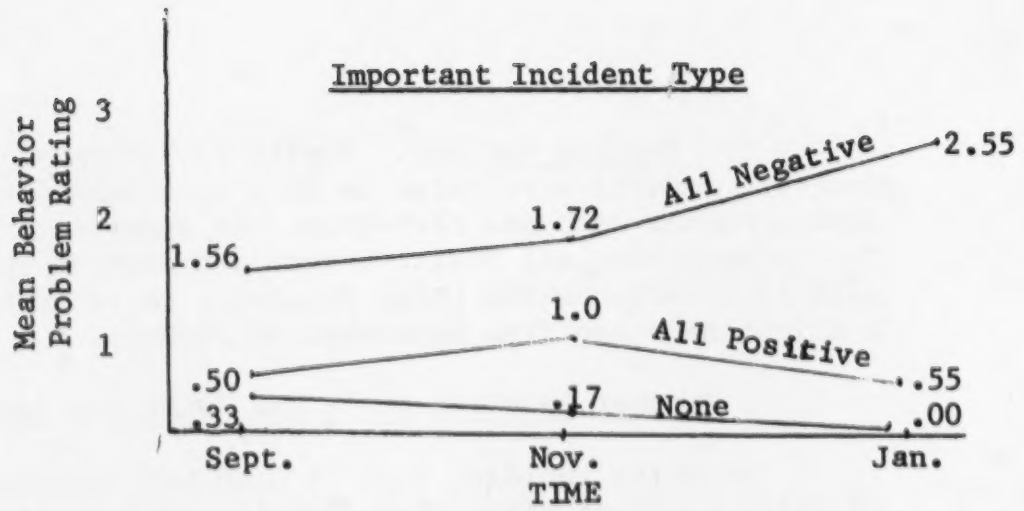


Figure 7.10. Behavior Problem Ratings: Important Incident type x time interaction.

on a documentation of the relationship between Incidents Reports and the more abstract ratings. The following significant effects were found:

1) Self-Anchoring Scale Ratings. Pupils receiving no Incidents at all and those receiving all Positive Incidents were rated relatively high by those teachers who gave primarily Negative Incident reports; teachers who gave mostly Positive Incidents rated these pupils much lower. (Figure 7.11). Conversely, pupils receiving all Negative Incident reports were rated as the lowest group by the teachers giving primarily negative reports, but were rated as the highest group by teachers who gave mostly positive reports.

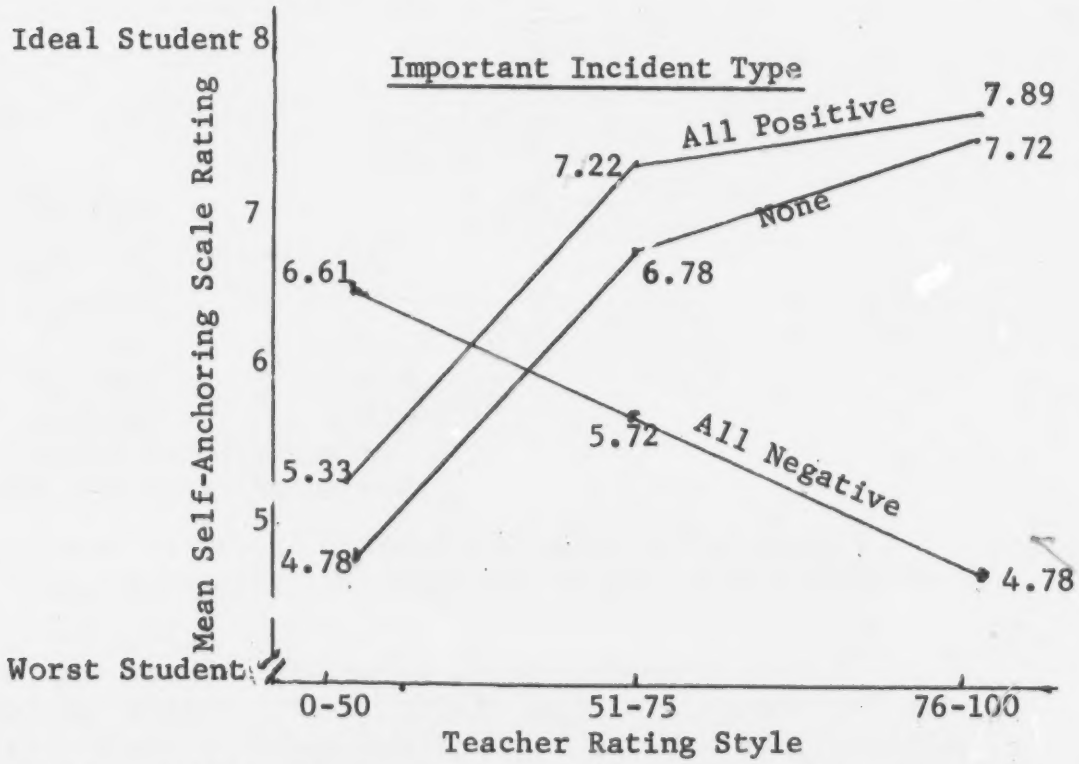


Figure 7.11. Self-Anchoring Scale Ratings: Important Incident type x teacher rating style interaction.

2) Problem Ratings. As described in the previous section, those pupils rated as high on Behavior Problems in September received all negative incidents throughout the semester and received increasingly higher Behavior Problem ratings. Pupils receiving no Incidents or all Positive Incidents were rated as low as Behavior Problems and remained low throughout the semester. (Figure 7.10).

It is interesting to note that not all pupils who had values discrepant from the teacher's (Figure 7.12) are rated as Behavior Problems. It appears that some children "act out" their values (i.e., the "all negative" group) more than others and thus become targets for the behavior problem label. (The results remained unchanged when Social Desirability scores were used as a possible moderator variable for the value measures.)

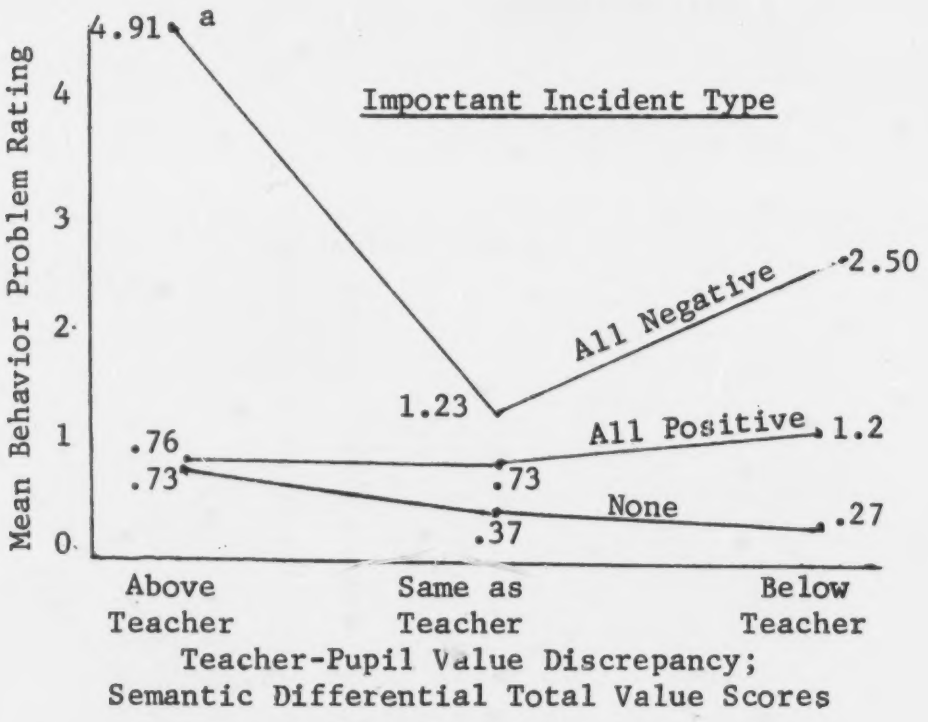


Figure 7.12. Behavior Problem Ratings: Important Incident type x teacher-pupil value discrepancy interaction.

[a]A Race X Teacher-Pupil Value Discrepancy Analysis of Variance for the "All-Negative" group confirmed this effect for both race groups. Analyses of children's Social Desirability Scale scores indicated that this was not a social desirability phenomenon.

Discussion

The procedures and results of this chapter have been presented under four major headings: Structural Analyses; Group Differences; Perceptual Shifts; and Teacher Labeling Style and Important Incidents. A brief discussion of the results included in each of these perspectives follows.

Structural Analyses

Factor analyses of teacher ratings revealed the presence of a generalized perception-factor defined by the Self-Anchoring Scale and had high loadings by the eight value dimensions as well as by the Personality and Behavior Problems ratings. Value Conflicts and Grade Point Average were also found to load moderately high on this general factor. These latter results suggest that most value conflicts, regardless of their specific value characteristics, are closely related to a global/evaluative perception of the pupil and also that this global factor, defined by perceptions of social adjustment, seems to include perceptions of pupil academic competence. It is interesting to note in this regard that the highest loading on the general factor belongs to the value "Academic Achievement," an appropriate loading given the setting of the research.

Group Differences

Results of analyses on the differential performance of pupils with varying backgrounds (i.e., grouped by social class, sex, and race) confirmed traditional hypotheses regarding the differential perceptions of certain social class and race groups. Specifically, middle-class pupils, females, and whites were rated more favorably than their counterparts whether the ratings were on the Self-Anchoring Scale, Personality or Behavioral Problems. One exception to this general effect was that middle-class blacks received the highest Behavior Problem ratings of any subgroup. These ratings tend to agree with the Value Socialization Scale mean differences but not with the Semantic Differential Total Value Score differences between groups.

Perceptual Shifts

The general pattern of shifts over time in ratings was puzzling. We had expected primarily monotonic functions which would help identify when and if teacher perceptions of pupils tend to stabilize or help confirm the existence of a cumulative "labeling" effect. One "labeling" effect did appear and is described in the next section. Otherwise, the findings generally revealed a marked "dip" in November with the September and January ratings similar and relatively high. Two plausible hypotheses which could explain these results are:

Hypothesis 1. The middle of the semester is essentially a period where perceptions are relatively unstable. In this epoch teachers may not be striving for a comprehensive/evaluative perception but responding to day-to-day fluctuations in transactions with pupils. This hypothesis is supported by the seemingly random appearance of the "dip" phenomenon across measures and across groups.

Hypothesis 2. The September and January ratings represent what could be termed a "Hello-Goodbye" phenomenon. September ratings were high because the teachers were anxious to cooperate but did not want to unduly criticize any child, thus, granting them the benefit of the doubt. January ratings were high because the teachers were anxious to complete the study and wished their class to "look good."

Both hypotheses could be true.

There was no evidence of major shifts in labels for any group over the semester.

Teacher Labeling Style and Important Incidents

Teachers giving primarily Negative Incident reports showed a range of expected behavior that was similar for most pupils; that is, Negative Incident reports were consistently related to lower global/evaluative ratings. However, teachers reporting mostly Positive Incidents seemed to individualize their expectancies more. Specifically, these teachers appeared to expect more from their most liked students (those with high global/evaluative ratings) while occasionally censoring them (giving negative reports) to keep up the high standards. On the other hand, these same teachers seem more tolerant of their less ideal pupils while primarily reinforcing them (giving positive reports) for improved behavior.

In general, those pupils who are rated as the most severe behavior problems in September receive all Negative Incidents and receive increasingly higher Behavior Problem ratings throughout the semester. This could be the result of a labeling effect per se. That is, once a pupil has been labeled as having a behavior problem, more of his specific behaviors are perceived as reinforcing the appropriateness of the label.

Naturally, these conclusions are all regarding teachers' reports. Some accuracy can probably be assumed for such reports but the sources of bias are unknown. For example, are the social class, sex, and race differences on these measures reflective of actual differences or of stereotypic distortion? The project collected no information that could answer this question.

CHAPTER VIII

VALUE CONFLICTS AND OTHER OUTCOMES

Values and Classroom Incidents

Throughout Study III's semester of data collection, teachers reported any important incident which occurred in their classrooms. It is of some interest to determine how many of these incidents were value related and how many were concerned with other matters (e.g., a child's personal problem). As noted earlier, the incidents were rated for their positive or negative nature and placed in one of the eight value categories or an "other" category.

A five way analysis of variance was used to determine whether or not the different types of incidents occurred with different rates. The three between factors were social class, sex, and race. Social class of the child was defined by his school's reputation. The first within group factor differentiated between incidents which were value related and incidents which were not value related. The second within dimension differentiated between positive incidents and negative incidents.

The race effects were insignificant at all levels of this analysis. All of the other factors interacted. The results of this analysis are presented graphically in Figure 8.1. This figure indicates that three-fourths of all classroom incidents were value related.

Figure 8.1 also shows that (1) Many more negative incidents were reported than positive incidents. (2) Males were more involved in both positive and negative incidents than were females. (3) The prime source of the interaction appears to be the lower class males who were involved in negative value incidents at a disproportionate rate.

The strikingly large percent of incidents involving values could be an overestimate. Due to the scales that the teachers filled out, they knew we were interested in values and so they may have reported more of their value incidents. Nevertheless, this overestimate would probably not be large: the teachers started reporting incidents before the study started so that feedback could be given on the type of report needed during the study. This feedback was solely concerned with the amount of information and not with the type of incidents.

The Occurrence of Value Conflicts

Group Differences in Value Conflicts

To determine whether particular types of children were

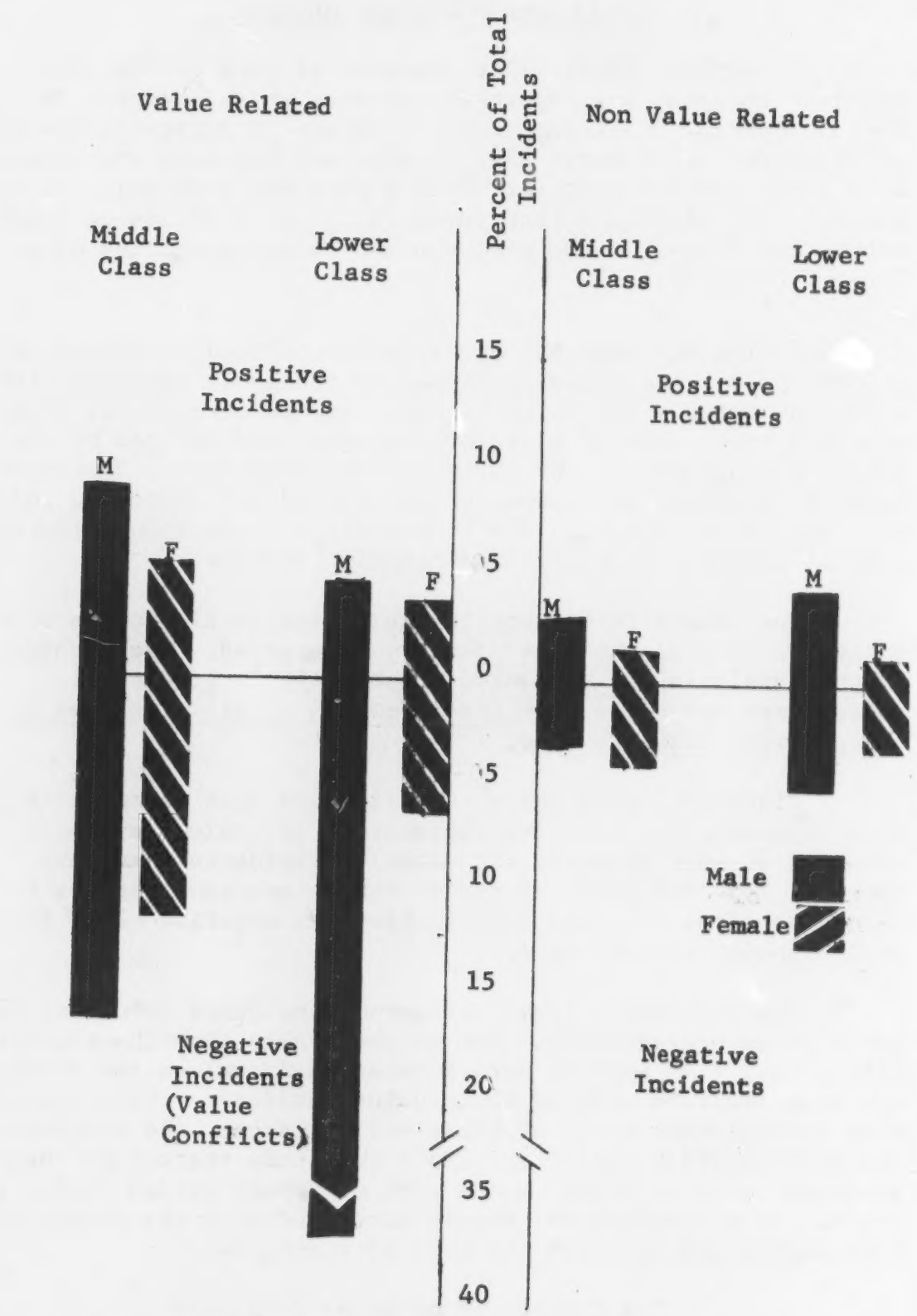


Figure 8.1. Value-related and non-value related incidents by social class and sex.

involved in more value incidents, another five way analysis of variance was performed. The three between dimensions were social class, sex, and race while the first within groups dimension was positive or negative incident and the second was the value category.

There was again no significant effect which involved race. Blacks and whites were involved in the same number of incidents as would be predicted from their class by sex subgroup. This in itself is a major finding since it contradicts common folklore and the teacher's perceptions of who are behavior problems (cf. Figure 7.2).

Figure 8.2 portrays the significant interaction which was found between social class, sex, the positive and negative nature of the incident, and the value categories. Considering the communality of occurrence across both social classes, the major category for value conflicts was Good Manners. Sex differences are readily apparent with the males being involved in more value conflicts in practically every category.

A striking feature of the differences across social class lies in the area of Diligence where the middle-class males were involved in many more positive incidents than were the lower-class males. The primary group involved in value conflicts over honesty were the lower-class males.

It is apparent that social class and sex were associated with differences in the types of incidents in which the children were involved. The interpretation of this finding is, however, more problematic. It could be a function of differential teacher expectations, of sex role training, of the values which each group holds and so forth. The present data is insufficient to examine such possibilities except as commented on below.

Value Conflicts As A Function of Values and Value Discrepancies

Value conflicts were designated as the criterion for a series of regression analyses. Each of these regression analyses, as well as the other regression analyses described in this chapter, used the measures collected at the beginning of the semester to predict later events during the semester. In this particular case, the prediction was to the total number of Value Conflicts in which the child was involved during the semester. These regression analyses first tested whether knowledge of group membership added to the prediction over and above the other variables, the effects of values in addition to the other predictors, and finally for the effects of the other personality variables. A general description of the procedure used is given in Appendix H. Since all social isolates had to be eliminated because they had no peer groups, the number of subjects for these analyses was 355.

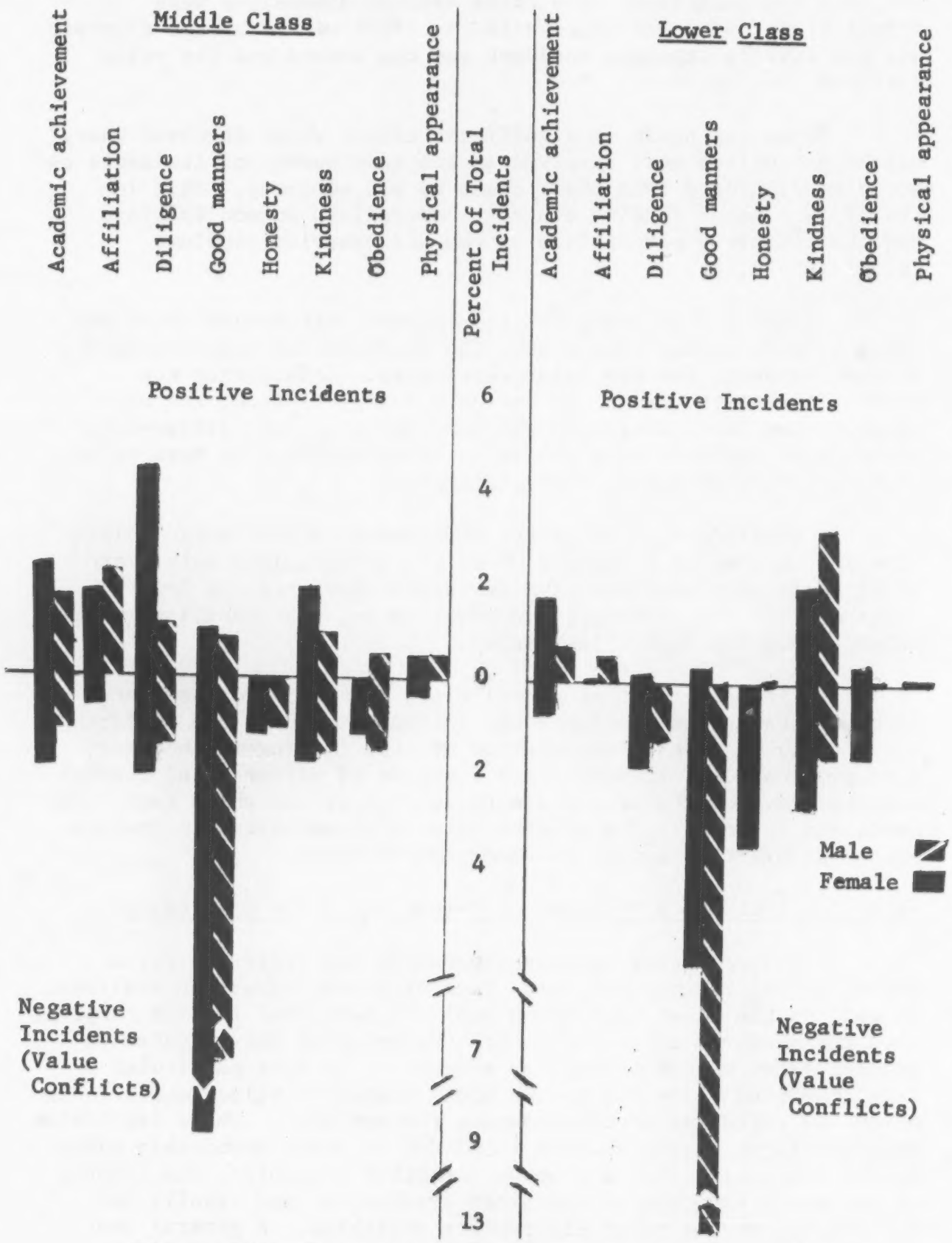


Figure 8.2. Percent of value incidents by sex and social class (school).

The first significance test was whether or not the total set of variables would predict value conflicts at a better than chance level. This they did and therefore the next test was run. This test found that it was indeed important to allow the variables to have separate weights in each of the eight groups formed by crossing social class, sex, and race. The null hypothesis that variables functioned in the same manner for all the groups was rejected. Further examination indicated that Extraversion, Non-Verbal Intelligence,the teacher's perceived Potency, and the Children's Social Desirability Scale all functioned differently depending upon which group the child was in.

The beta weights predicting value conflicts from only the significant measures are given in Table 8.1. In that table, Extraversion, Potency, and Social Desirability operated in the expected direction for most of the groups. The more extraverted male, the child who saw the teacher as weak, and the child not concerned with being socially desirable had the greatest number of value conflicts. Intelligence operated in a more ambiguous manner since its weight was sometimes positive and sometimes negative. In the former case the weight suggested that the brighter child was involved in more value conflicts but in the latter case the weight suggested that the brighter child was involved in fewer value conflicts.

Teacher-pupil discrepancies on the Semantic Differential Total Value Scale predicted value conflicts in the same manner in all groups. A positive difference, i.e., a difference which showed the teacher to be above the pupil, was associated with more value conflicts. However, the weight given this discrepancy score was quite small.

The pupil-peer discrepancies on the Value Socialization Scale had a complex relationship to value conflicts. Not only was there a significant social class by sex by race by scale interaction, the relationship was also significantly curvilinear. The interpretation of beta weights when the relationship is linear is rather straightforward compared to the case of curvilinearity. To better understand the relationships, predicted value conflicts were estimated for each of the eight subgroups. This estimation assumed the child had an average score on each of the other predictors and varied only on the discrepancy scores.

An examination of these scores suggested that caution should be exerted in directly interpreting the weights. For example, the black middle-class males had scores which suggested that the contribution of the discrepancy scores to their value conflict prediction was negligible despite relatively large--but offsetting--weights. On the other hand, the white lower-class males had predicted value conflicts which varied more widely than any other group.

Table 8.1

Beta Weights for Predicting
Value Conflicts

	Race	Groups							
		Blacks				Whites			
	Sex	Males		Females		Males		Females	
Variables	SES	Lower	Middle	Lower	Middle	Lower	Middle	Lower	Middle
Group Vectors		- .55	.00	.00	.00	.00	-.01	.35	.00
Extraversion		3.57	.49	.00	.00	.71	.00	.00	-.37
I Q (non-verbal)		.32	.03	.07	.04	-.12	-.07	.01	-.01
Potency		.27	-.46	-.05	-.19	-.95	.07	.35	.19
Social Desirability		-3.13	.00	-.72	.00	.34	-.15	.00	.12
Semantic Differential Total Value Score Discrepancy from Teacher		← .15 →							
Value Socializaton Scale: Discrepancy from Peers		.16	.04	.06	.02	.07	.01	-.09	.02
Discrepancy from Peers Squared		.34	.48	.01	-.01	-.05	.06	-.14	.00
Discrepancy from Peers Cubed		- .36	-.34	-.05	.00	.15	-.01	-.08	.03

R = .71; Unbiased (shrunken) estimate of R = .62.
N = 355.

An examination of the Value Conflicts predicted from the peer discrepancies suggested that the interaction was indeed complex. For some groups, such as the black middle-class males and white lower-class males, the relationship was curvilinear. Any child who differed from his peer group was more likely to be involved in value conflicts. The relationship went in the opposite direction for the white lower-class females with anyone deviating from their peer group being less likely to be involved in value conflicts. As if this were not complicated enough, the black lower-class females, for example, had essentially a linear relationship with those below their peers being involved in more value conflicts while those above their peers were involved in less conflicts on the average. The most consistent results appeared to be that pupils below their peers had generally more value conflicts.

The prediction of value conflicts was good with a multiple correlation of .71. Even the multiple correlation corrected for the possibility of capitalizing upon chance inherent in such an equation was reasonably high (.62). These variables could be used to suggest which children might have more value conflicts in the school setting so that aid could be offered.

Other Outcomes as a Function of Values and Value Discrepancies

Anxiety

It is possible that one outcome of value discrepancies would be an increased anxiety level in the child. This should be particularly noticeable around the time of the incident as measured by a measure of state anxiety but, if enough such incidents occurred, it is reasonable to expect it to affect the trait level of anxiety as well. Therefore the same set of analyses by which the personality and value measures were tested for predicting the value conflicts was also used to test for the prediction of Anxiety at the time of the second testing session.

In these analyses, Anxiety scores taken at the first of the semester were carried as a covariate. This allowed for an analysis where the attempt was to predict that part of the second Anxiety measure which was not predictable from the first Anxiety measure. This approach to change avoids some of the problems of difference scores (Cronbach & Furby, 1970) even though it has its own problems.

The overall test of the predictability of the second set of Anxiety scores was significant but the interactions of both non-value and value measures with the groups did not increase the prediction achieved without the interactions. Neither were the group main effects significant. When analyses were computed to determine whether or not values in linear, quadratic, or cubic form predicted the second anxiety measure over and above the first Anxiety measure, it was found that they did not. Further checks on the other variables led to a model which contained Extraversion, Intelligence, and Potency in addition to the first anxiety measure.

The three predictors received negative weights (cf. Table 8.2) which suggested that the introverted, less bright child who viewed the teacher as relatively weak was more likely to increase his score on Anxiety during the semester. This is understandable since an introverted, dull child may need the more structured situation provided by a strong teacher.

Table 8.2

Beta Weights for Predicting
Mid-Year Anxiety

Variable	Weight
Anxiety, Fall (covariate)	.43
Extraversion	-.10
I Q (non-verbal)	-.09
Potency	-.10

R = .49; unbiased (shrunken) estimate of R = .49.
N = 355.

Grades

Each child's Grade Point Average in those subjects taught by the teacher doing the ratings formed another criterion. While this measure is not completely free from the teacher's biases, most teachers do base part of their grades on objective data. The overall prediction of grades from the pool of predictors was significant but the interactions for the non-value and value measures were nonsignificant. Neither did group main effects significantly increase the prediction over that achieved by the personality and value measures alone.

Table 8.3 presents the beta weights for the final model to predict Grade Point Average. The Anxiety and Intelligence weights are directly interpretable; Social Desirability's negative weight in predicting grades is not so easily understood. Perhaps the child who is concerned with social desirability is more interpersonally and less achievement oriented.

The teacher-pupil discrepancies on the Semantic Differential Total Value Scale did have a curvilinear relationship with grades. This relationship is plotted in Figure 8.3. This graph indicates the difference in grades which would be predicted with different value scores. For example, a child's being 40 Semantic Differential points above the teacher would lead to a predicted Grade Point Average only .3 greater than if he had the same score as the teacher. Being 40 points below the teacher would lead to more change in the predicted score, i.e., a drop of .5 in the child's Grade Point Average from what it would have been if he were the same as the teacher.

Table 8.3

Beta Weights for Predicting
Grade Point Average

Variable,	Weight
Anxiety	-.16
I Q (non-verbal)	.44
Social Desirability	-.10
Semantic Differential Total Value Scale:	
Difference from Teacher	-.30
Difference Squared	.10
Value Socialization Scale:	
Child's Score	.28
Difference from Peers	-.30

R = .62; unbiased (shrunken) estimate of population R = .61.
N=355.

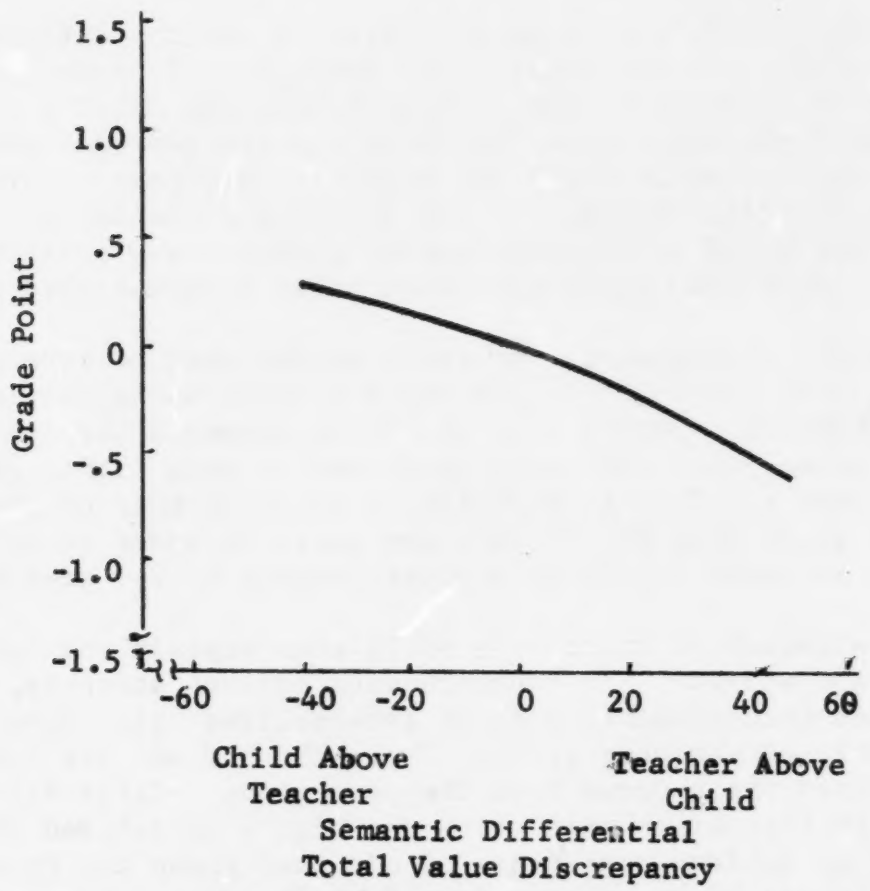


Figure 8.3. Predicting grades from teacher-pupil discrepancy on the Semantic Differential Total Value Scores.

Figure 8.3 indicates that the major alteration in predicted Grade Point Average occurred when the teacher was definitely above the pupil. This value discrepancy appeared to disrupt the learning process. The student's being above the teacher had less decisive affect upon the Grade Point Average but was associated with somewhat higher grades.

In interpreting Figure 8.3, it is important to note that the child's own Semantic Differential Total Value Score did not add significantly to the prediction. This means that it was not the student's portion of the discrepancy score which was predictive. If that were the case,then the purer measure of that score, pupil's score itself, would have been the better predictor and the discrepancy score would not have added significantly to it. But the discrepancy score did.

The lack of the pupil's value score being included in the model also indicates that it was not the teacher's score alone that added to the prediction. If this were the case, the pupil's value score would have been entered into the equation to suppress that part of the discrepancy score which came from the pupil. Rather, this beta weight could have arisen only because the difference between the two people involved in the situation is more important than either's absolute score on the value scale.

On the Value Socialization Scale, it was the difference from peers which was the significant indicator of grade point average. Its negative weight indicated that the child's Grade Point Average was lower when the child's peers scored lower on the Value Socialization Scale and higher if his peers scored higher than he did. Coleman (1966) observed a similar phenomena: a lower-class child in a predominately middle-class classroom performed higher than if he were with other disadvantaged children.

But the discrepancy from peers weight must be interpreted in light of the loading from the child's Value Socialization scale itself. This had a positive weight which suggests that as the child was more socialized, then he tended to make higher grades. It may be that the crucial variable is actually that of the peers. If that is true, then the child's own score is added in as a corrective in order to obtain a purer measure of the peer score.

An alternative hypothesis could also explain the two Value Socialization weights. In research with college students, Bowers (1968) found that students with an internalized value functioned independently of the peer group. Those who had not yet internalized a value gained their norms from the peer group. Table 8.3 is consistent with this hypothesis since the highly socialized child would tend to achieve regardless of his peer group and thus his scores would add to the prediction of his grade point average.

The child who is less well socialized will be more dependent upon his peers for his norms of behavior, and therefore obtain lower grades if they were less socialized and higher grades if they were more socialized. Since both effects would be operating, both variables would have significant beta weights. The extent to which one effect could override the other is not known.

The inclusion of value and personality predictors increased the percent of predicted variance in grades by 50% over that predictable from intelligence alone. More reliable measures of intelligence, personality and values could be expected to increase the multiple correlation still more.

The Prediction of Teacher Perceptions

Personality and Behavior Problem ratings were chosen to represent the teacher perceptions. They both could predict the general perception factor by being added together and yet provided independent information while reflecting behavior crucial to the classroom setting. Since these are subjective perceptions, it would be expected that the inter-personal relationship between the teacher and the pupil would be of more crucial importance. This was found to be true since a value discrepancy measure added to the prediction of both personality and behavior problems. No peer discrepancy measure nor any child's value scores alone added to the prediction in either case.

Personality Problem Ratings

The total equation containing all predictors in interaction with all the groups was significant. Further analyses indicated that the Social Desirability Scale and Intelligence both interacted with group membership. The interactions of the groups with one value scale was also significant. Both teacher discrepancy measures (i.e., on the Semantic Differential Value Scale and on the Value Socialization Scale) had curvilinear relationships to Personality Problem ratings.

The final model is summarized in Table 8.4. Intelligence contributed to the prediction within most groups with the general trend being towards decreased problems with higher intelligence. The Social Desirability measure was negative for the black middle-class males.

The curvilinear relationship between the teacher-pupil discrepancies on the Semantic Differential Total Value Scores and the Personality Problem rating is plotted in Figure 8.4. This shows that the major predictability occurred when the pupil was above the teacher. As the pupil scored higher compared to the teacher, the teacher was less likely to see him as having personality problems. The prediction did not operate in the opposite direction; when the

Table 8.4

Beta Weights for Predicting
Personality Problems

	Race	Groups: Blacks				Whites			
	Sex	Males		Females		Males		Females	
Variables	SES	Lower	Middle	Lower	Middle	Lower	Middle	Lower	Middle
Group Vectors		.26	1.78	.01	.51	.00	.00	.11	.06
I Q (non-verbal)		.06	- .09	-.06	.10	.10	-.15	-.13	-.18
Social Desirability		.14	-1.57	.00	-.30	.07	.00	-.05	.00
Semantic Differential Total Value Score:									
Discrepancy from Teacher		←		.13		→			
Discrepancy from Teacher Squared		←		-.16		→			
Value Socialization Scale:									
Discrepancy from Teacher		-.49	-.47	.08	-.30	.08	.03	-.04	.01
Discrepancy from Teacher Squared		.34	.34	-.09	.14	.04	.01	.06	-.01

R = .61; unbiased estimate of population R = .54.
N = 355.

teacher was above the pupil, little differential prediction could be made.

Without additional data, it is not known whether or not the teacher's perception is accurate. Perhaps these children do actually have fewer personality problems, an interpretation compatible with Mowrer's (1964) and London's (1964) theories and supported by Peterson (1967), or it could be that such a child is "separated out to be ignored." In the latter case, the teacher may pay closer attention to the children who have values lower than herself and thus be more sensitive to their unique problems.

To help understand the four way interaction of groups with teacher-pupil discrepancy on the Value Socialization Scale, predicted Personality Problem ratings were made for variations in

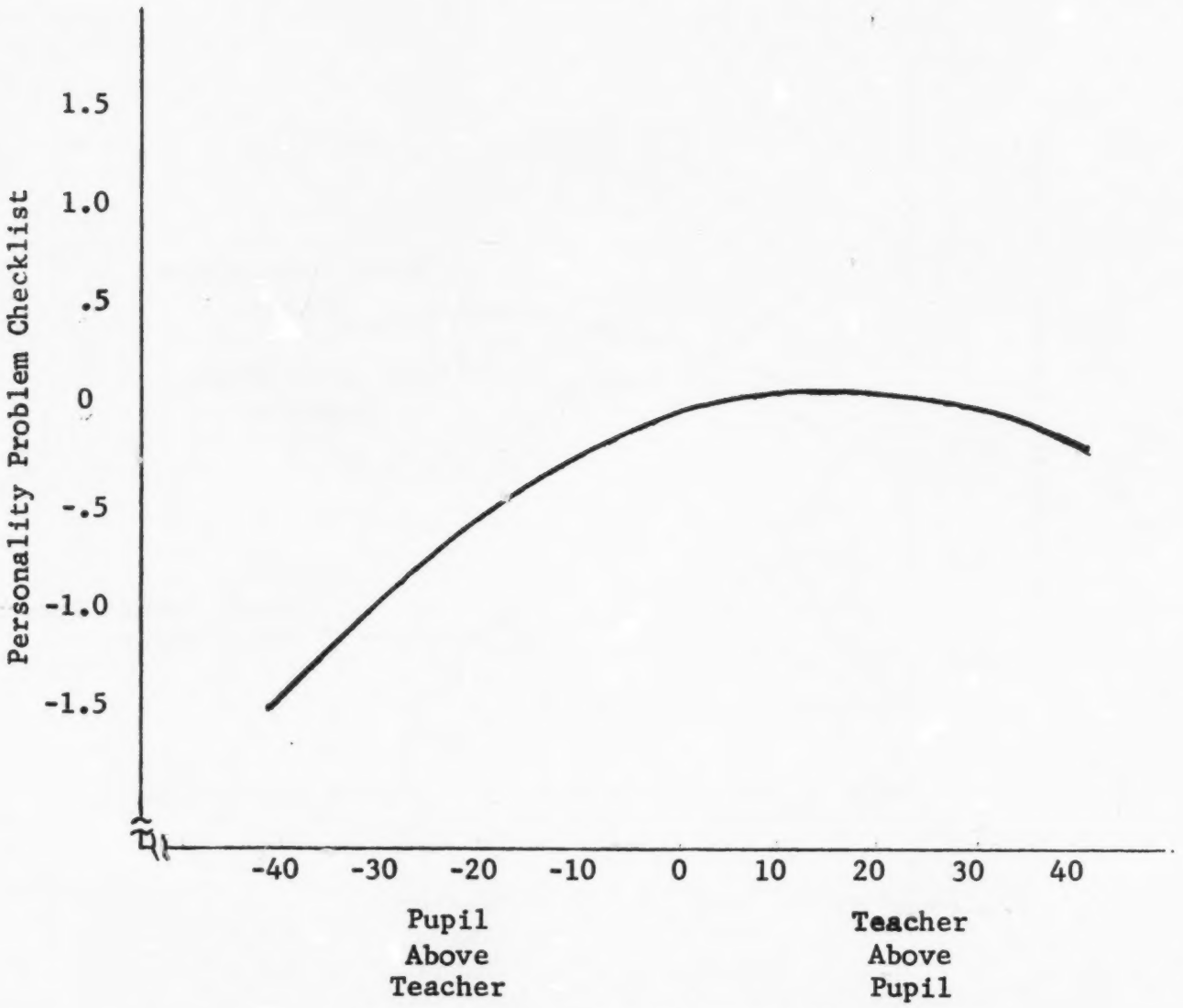


Figure 8.4. Predicting Personality Problem ratings from discrepancies on the Semantic Differential Total Value Scores.

that scale while the other variables were held constant. Three of the typical curves which were generated by the procedure were plotted in Figure 8.5.

Only the range of predictor scores which included the observed range for that group were plotted in figures such as 8.5. The differences in levels of the curves indicate the relationship of this predictor to the set of other predictors in this context. If the other predictors' weights overestimate the Personality Problem score, then the curve would indicate that

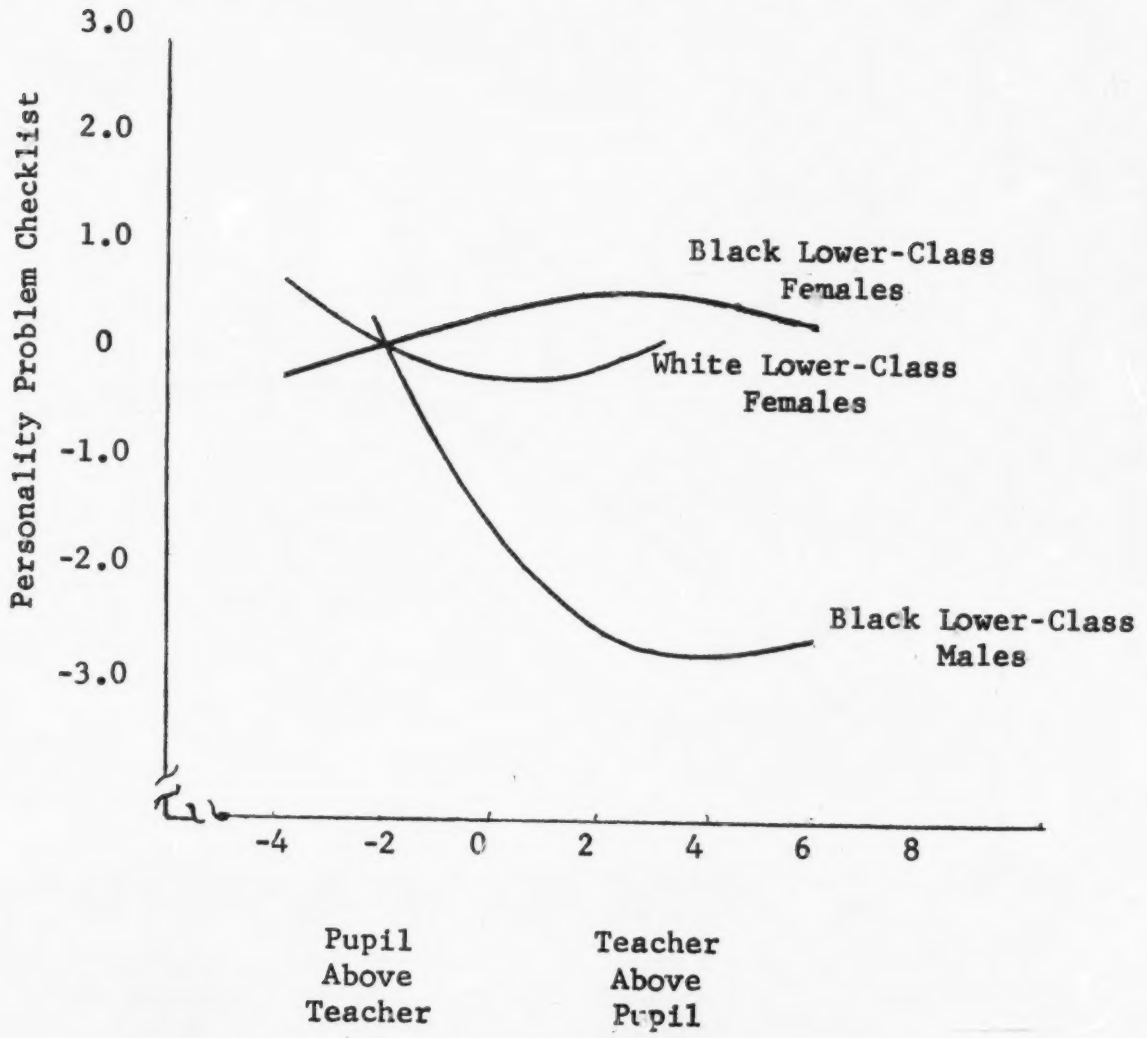


Figure 8.5. Relationship of teacher-pupil Value Socialization discrepancies to Personality Problem Rating.

Note: These three groups were selected as arbitrary samples of the type of relationships found. Only the range for which scores actually occurred is plotted.

Personality Problem points should be subtracted in varying degree for different scores on the predictor. If the other variables' weights lead to an underestimation of the Personality Problem ratings without this predictor, then the curve would indicate how many points should be added to the ratings to increase their accuracy. For our purposes, the mean level differences between the curves is less important than the type relationship indicated by the curve.

Figure 8.5 illustrates two features of the relationship between these discrepancy scores and Personality Problen ratings. First, the curves follow radically different forms. The black lower-class female curve is the reverse of the white lover-class female curve, and the lower-class black males have a more pronounced curve than either of the others.

Second, the curves leading to greater differences on the Personality Problem checklist tend to be opposite in form to the curve given in Figure 8.4. The pupil who was above the teacher tended to receive fewer Personality Problem checks than did the child who is below the teacher on the Semantic Differential Total Value Score. But to receive a minimum rating on the Personality Problem checklist as a function of Value Socialization, it appears that the child must be three to five points below the teacher since the curves tend to break in this area. (These two different forms of curves are of course possible since the two different value measures were correlated at only a low level.)

If the attitude scale is seen as reflecting feelings and the Value Socialization Scale is viewed as reflecting "oughts," then it appears that a more positive attitude towards socialization values was viewed by the teacher as indicative of a sound child but that admiration of such a child was not.

Since the Value Socialization Scale does interact with the group to which the pupil belongs, teachers may have different implicit norms for what a child should admire if he is a member of a particular group. According to this interpretation, Figure 8.5 suggests that the child who surprised the teacher by going against the stereotype was seen as most likely to have a personality problem. For example, only black lower-class males who were actually below the teacher on socialization tended to be seen as having few Personality Problems. Those as high as their teachers were candidates for a higher rating. Perhaps this is because the child below the teacher is "keeping in his place." The teacher may have assumed that no child in this group could be well socialized and considered any such behaviors as evidence of some disturbance.

The level of the multiple correlation suggests that some prediction of the child being viewed as having personality problems can be made even though the complex interactions are not directly interpretable from present theory.

It should be noted that anxiety, the prime personality characteristic associated with a neurotic diagnosis, is not related to the teacher's view of the child as having personality problems. Evidently, the child is equally happy or unhappy regardless of whether or not the teacher thinks he has problems.

A suspicion that part of the complex interaction of the Value Socialization Scale with race and other group membership variables is a function of differences between teachers in different schools is probably not supportable. The teachers across all the schools received the same average score on this scale. The scale on which the teachers differed, the Semantic Differential Total Value Score, was where the Personality Problem prediction followed the same form across groups.

Behavior Problem Ratings

The overall prediction of Behavior Problem ratings was significant and the group interactions with Extraversion were also significant. The Children's Social Desirability Scale was also significant but it did not interact with the groups. Analyses of the values indicated that there was only one curvilinear relationship but that the values did interact with the groups. Further analysis indicated that the four-way interactions of social class by sex by race by value scale were not significant. The teacher-pupil discrepancy on the Value Socialization Scale did have a significant two-way interaction with sex and social class. In addition, there was also a race by discrepancy interaction for the same scale.

Table 8.5 gives the beta weights from the final solution.

Table 8.5

Beta Weights for Predicting
Behavior Problems

	Race	Groups							
		Blacks				Whites			
	Sex	Males		Females		Males		Females	
Variables	SES	Lower	Middle	Lower	Middle	Lower	Middle	Lower	Middle
Group Vectors		.42	-2.18	.12	-.07	-.07	.29	.09	.16
Extraversion		.00	2.46	.00	.18	.17	-.30	-.06	-.21
Social Desirability		←			-.15				→
Value Socialization Scale:		Males			Females				
		Lower	Middle		Lower	Middle		Blacks	Whites
Discrepancy from Teacher		-.01	.04		.11	.02		-.52	.03
Discrepancy from Teacher Squared								.38	

R = .53; unbiased (shrunken) estimate of population R = .49.

Whereas it had been predicted that Extraversion would be important in producing behavior problems, this appeared to be the case only within certain groups. In particular, the black middle-class males tended to have more ratings as a behavior problem if they were extroverts than did others. Social Desirability had a consistent relationship, with the child with more social desirability motivation being less involved in behavior problems, but the relationship was small.

The sex by class interaction on the Value Socialization Scale is directly interpretable from its weights. It would seem that there is a slight tendency for lower-class females with Value Socialization Scale scores lower than their teacher to be seen as having more problems. Although this interaction was significant, it appears that its effect was not great.

The other interaction with the Value Socialization discrepancy from teacher measure is plotted in Figure 8.6.

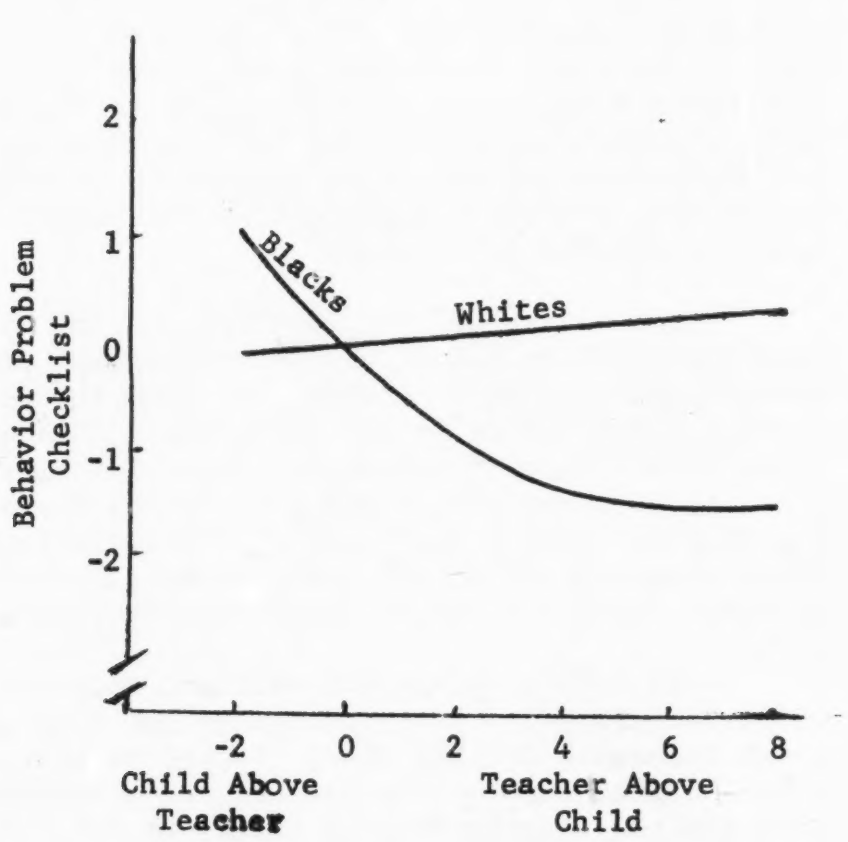


Figure 8.6. Race differences in predicting Behavior Problems from teacher-pupil Value Socialization discrepancies.

As this plot indicates, the whites had a trivial linear prediction from their discrepancy scores to their Behavior Problem ratings.

On the other hand, those blacks who were above the teacher on the Value Socialization Scale were viewed as being more of a behavior problem with those lower than the teacher not receiving such ratings. This result is consistent with one of the interpretations suggested for the group by Value Socialization interaction predicting Personality Problems. It also seems reasonable here that those blacks who surprised their teacher by being as much as or more concerned with values than the teacher herself have difficulties. Perhaps the teacher feels they are not being honest with her or perhaps this threatens her position as the leader of the classroom.

The overall prediction of Behavior Problem ratings from this set of variables is approximately as good as one generally finds from intelligence to grades. The multiple correlation was .53, with an unbiased (shrunken) estimate of the population correlation being .49.

Values and Value Discrepancies as Predictive Constructs

It does appear that values and value discrepancies predicted classroom behavior as measured by the current procecures. For value conflicts, grades, and personality problems, both value scales added significantly to the prediction. This would suggest that developing value categories from children's own statements led to values related to the classroom. Only in the case of anxiety was there no relationship between the value measures and the criterion.

Most of the value prediction was not from the child's absolute stand on the values but rather, as predicted from a social psychological perspective, from the discrepancy in values between participants in the situation. This was partially a function of the criteria used. Most of the criteria reflected the interpersonal events going on in the classroom or the perceptions which were developed by the socializing agent. Since these characteristics were interpersonally oriented, an interpersonal aspect of values would be more relevant.

The complex group interactions occasionally found with the Value Socialization Scale cannot be put into an overall theoretical framework at this time. It may be that the different groups have somewhat distinctive manners of interacting which tend to moderate the relationship of values to the criteria. Or it may be that teachers have differential expectations for these groups and these expectations produce the effects. Other interpretations are also possible.

The curvilinear prediction occasionally found suggested that tolerance limits probably exist. Figures 8.5 and 8.6 suggest

that the tolerance limits were lower for some who were higher on the Value Socialization Scale than they were for children who were lower on the scale. Perhaps teachers expect young children to be less socialized and therefore accept this, but do not like the "put down" effect which arises when the child is "better" than she. On the other hand, Figure 8.4 suggests an acceptance norm whereby a child is more accepted as he moves above the teacher on the Semantic Differential Total Value Score.

We wish we could give a definitive statement of the differences between the two value measures used in this project at this point. Such a statement would increase our understanding of why they often operate in different ways. However, it is felt to be too early to attempt theoretical descriptions of how these scales differ. At this point, it seems more appropriate to simply recommend that we carry both in future research so that a larger body of data can be recorded as a basis for an interpretation.

The magnitude of the contribution of values to predicting the criteria was not particularly great even though it was significant. The increase in predictability gained by adding a significant value measure to the equation was only a few percent of the total variance. Several reasons can be suggested for the lack of higher correlations. (1) Criteria such as grades and value conflicts are not noted for their reliability and the extent to which the criterion is unreliable will limit the prediction from any variable. (2) The present value measures were newly developed scales and had reasonable reliabilities for research purposes but not sufficient for use with individual subjects; only the latter type of reliability could be expected to account for large shares of variance. (3) Criteria generally had skewed scores which tended to reduce the possibility of prediction. (4) The bias analysis suggested that some of the more extreme pupils were excluded from the study.

A possible major restriction on the prediction from values and value discrepancies was the changes in values that occurred. This was predicted from a transactional perspective and supported by the data. However, statistical models are not yet transactional and the adaptation effects could not be taken into account. Instead, the initial measures had the double task of predicting the criteria while implicitly taking into account the ongoing transaction.

Another major source of limitation on the prediction of the criteria by the value measures probably arises out of the nature of situation. Due to the highly structured nature of the classroom, little room is left to the child to make his own decisions. It has been shown (Schwartz, 1968) that values relate

highly to a person's behavior only when that person feels some degree of responsibility in the situation. Unfortunately, the moderator variable we included as a rough measure of that, a children's locus of control scale, failed to be usable with children.

A further reason for the lack of predictability is that there are many ways of resolving value differences which do not necessarily change the values nor lead to changes in the type of criteria examined here. Indeed, it could be argued that the best teacher is the one who is able to help the child achieve high grades and stay out of classroom conflicts while not allowing their value differences to distort her perception of them. Such an "ideal teacher" would create a situation where values would then not predict behavior. She would probably also have minor effect on her children's value development.

The importance of these results is determined by their usefulness rather than the percent of variance accounted for. It is not expected that the current value measures will be used to predict overt conflicts in the classroom, grades, or the other criteria, a situation requiring the ability to account for large shares of the variance. Rather, the research study was undertaken as a field study to provide leads on the conditions under which values do influence the classroom events. The present significant results are therefore important because they provide points of departure for developing experimental programs which will tend to alter conditions in the classroom so as to produce a better classroom atmosphere. These could, for example, take the form of shifting teachers' expectations so that they will be more understanding of differential life styles and realize the basic agreement in values that all groups bring to the classroom.

Or the significant results could lead to experimental and quasi-experimental studies designed to aid the children directly. For example, if it is decided to help the child increase his value scores, then placing him with a teacher who is above him in values and cutting him off from peer group support at the lower value level could be seen as important. Such procedures would, hopefully, increase the child's performance in the classroom as shown by the prediction of grades.

CHAPTER IX

SUMMARY AND CONCLUSIONS

The Research Project

Values are generally felt to be important determinants of how a person responds in a given setting. Those people who enter the setting with an inappropriate set of values would be expected to be involved in a disproportionate share of conflicts. The school system is one particular setting which is entered by children with a wide range of values. To what extent might these discrepancies between the children's values and the values of the schools lead to problems? And to what extent are the value discrepancies related to subcultures identified by social class and race? This project sought to investigate these and related topics.

To investigate values, a measurement technique was needed. Such a procedure should be appropriate to all the children regardless of their subculture. This project defined values by going to the children themselves. The value categories which lower-class blacks, middle-class blacks, lower-class whites and middle-class whites used in response to open-ended inquiries were identified. All groups used the same value categories. These included: Academic Achievement, Affiliation, Diligence, Good Manners, Honesty, Kindness, Obedience, and Physical Appearance. The values were highly related; a child who held to one held to the other values as well. This interrelationship identified a basic value dimension of social conformity or socialization which was selected as the thrust of the value measurement. Two different types of questionnaires were used to measure total value socialization. The final questionnaires contained only items which had been validated within each of the eight subgroups formed by crossing social class (lower and middle) with race (black and white) and sex (boys and girls). This requirement increased the likelihood that the scales would be fair to each group.

Fourth and fifth grade children were given the value scales and other measures at the beginning of a fall term and in the following January. Each teacher rated her children and reported any important incidents which occurred in their classes. These incidents were defined as any events which stood out as the teacher recalled the events of the past several days. These events could be positive or negative. Eight schools--two from each of the categories formed by crossing race and social class--provided a fourth and a fifth grade classroom for the study. Pupils participated only if their parents agreed for them to do so.

When interpretating the conclusions reached, one should keep in mind the type of child involved, the way values were

measured, and the fact that the project was a field study. This means that the project does not allow recommendations to be made, for example, on how teachers could shift the values of ghetto children. Rather, the project provides for a further understanding of the role of values in more traditional school settings.

Teachers and Values

The teacher's impact as a socializing agent was relatively clear. Children who were lower than their teacher on either value scale at the beginning of the term shifted up on that scale as the term progressed. Those children who were initially above their teacher did not change. Children who had teachers who scored higher on the scale were, at mid-year, higher on the scales than those children whose teachers were not so high.

The score of the average child shifted across time so as to reduce value discrepancies between the children as a group and their teachers. By the middle of the fifth school year, the children had mean values about the same as the mean values of the adults. This appeared to be a major method of "adapting" to value discrepancy.

The teacher probably plays a greater role in value development than had been previously supposed. This is not surprising from Bronfenbrenner's analyses (1970). He pointed out that parents and other adults spend only an hour or two a day with children; the teacher normally spends four or more hours a day with those same children.

The Impact of Values

Values aided in predicting all the criteria except changes in anxiety. This prediction was not, however, from the child's values alone but generally from the discrepancies between his values and his teacher's or peer's values. Sometimes it was the child who scored above his teacher or peers who was the problem and sometimes it was the child who scored below. The form of the relationship usually suggested that value discrepancy was critical only after it reached a given level. These findings supported the social psychological position that the classroom is best viewed as an ongoing transaction among the teacher, child, and peers.

Increases on the values identified in this project may not be universally desired since the socialization dimension did have a social conformity component in addition to its component of concern for others. On the other hand, attaining values such as obedience may be important for this age child because he has to adapt to a conformity oriented adult system or because social conformity is a transitional phase necessary for movement to a more mature ethical system (Kohlberg, 1969). Social scientists cannot answer questions concerning the desirability of changing values, but our data do suggest that values have an impact on the

classroom and that the school setting influences the child's values. (Cf. Kohlberg, 1970, and Gorsuch, 1972, for further discussions of the implications of our current knowledge for education.)

Value Conflicts, Value Discrepancies and Subcultures

Over two-thirds of all the reported incidents involved values. The value conflicts, i.e., the negative incidents involving values, formed a major share of the incidents and were predicted by child-teacher differences on one of the value scales. Value conflicts were also related in a complex manner to child-peer differences on the other value scale.

The overt value conflicts were related to both social class and sex. Lower-class children and males were involved in more value conflicts. However, the teachers' perceptions of the children were somewhat different. They rated middle-class black males--and particularly those who were above the teacher on one of the value scales--as presenting more behavior problems even though race was unrelated to the number or type of incidents in which a child was involved.

The value discrepancies did not appear to be strongly related to the subcultures. While lower-class males were consistently lower on value measures than, for example, middle-class females, these differences were too small to be a major source of problems. The average child in each group had values compatible with the values of the average child of each of the other groups. Indeed, differences in value discrepancies across groups were also found to arise from differences between the teachers. Teachers in black schools scored lower than teachers in white schools on one of the value scales even though both sets of the teachers were predominantly white; the differences between these two sets of teachers were as great as the differences between children's groups on that scale.

The value scales occasionally functioned in a different manner for children from the different subcultures. It was not obvious whether this was a function of actual differences between the groups or whether it was a function of the teachers' holding different expectations for the different groups. Such expectations could produce different perceptions by the teacher and result in altered behavior by the children.

If negative classroom incidents are considered to be hinderances to the educational process, then attention to value conflicts is certainly necessary. Part of this attention may need to center on the further training of teachers, both to alert them to possibly misleading expectations of certain subcultures and to cope creatively with value laden incidents in their classes.

REFERENCES

Allport, G. W., Vernon, P. E., & Lindzey, G. A manual for the study of values. Boston: Houghton Mifflin Company, 1960.

Baker, H. L. A comparison of the personality characteristics of selected third grade pupils from three socioeconomic status levels. Unpublished Doctoral Dissertation, University of Toledo, 1966.

Barthol, R. P., & Bridge, R. G. The ECHO multiresponse method for surveying value and influence patterns in groups. Psychological Reports, 1968, 22, 1345-1354.

Barthol, R. P., & de Mille, R. Project ECHO final report. Santa Barbara, California: General Research Corporation, March, 1969.

Bialer, I. Conceptionalization of success and failure in mentally retarded and normal children. Journal of Personality, 1961, 29, 303-320.

Bills, R. E. The effects of a value on learning. Journal of Personality, 1952, 21, 217-222.

Bottenberg, R. A., & Ward, J. H., Jr. Applied Multiple Linear Regression. Technical Documentary Report, Lackland Air Force Base, Texas, 1963.

Bowers, W. J. Normative constrains on deviant behavior in the college context. Sociometry, 1963, 31, 370-385.

Bradford, D. L. The formation of achievement attitudes among lower class Negro youth. Unpublished Doctoral Dissertation, University of Michigan, 1966.

Bronfenbrenner, U. The role of age, sex, class, and culture in studies of moral development. Religious Education, 1962, 57, S3-17.

Bronfenbrenner, U. Two worlds of childhood. New York: Russell Sage, 1970.

Brozovich, R. W. Group norms in sixth grade classrooms of contrasting socioeconomic status and differing racial composition. Unpublished Doctoral Dissertation, George Peabody College, 1966.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Company, 1963.

Campbell, E. Q., The internalization of moral norms. Sociometry, 1964, 27, 391-412.

Cattell, R. B. Culture fair intelligence test. Champaign, Illinois: Institute for Personality and Ability Testing, 1950.

Cattell, R. B. Personality and motivation structure and measurement. New York: World Book Company, 1957.

Cheyney, A. B. Teachers of the culturally disadvantaged. Exceptional Children, 1966, 33, 83-88.

Cloward, R. A., & Jones, J. A. Social class: Educational attitudes and participation. In A. H. Passow (Ed.), Education in depressed areas. New York: Teachers College, Columbia University, 1963.

Cohen, E. G. A new approach to applied research: Race and education. Columbus, Ohio: Charles E. Merrill Publishing Company, 1970.

Cohen, J. Multiple regression as a general data analytic system. Psychological Bulletin, 1968, 70, 426-443.

Cole, C. W., & Miller, C. D. Relevance of expressed values to academic performance. Journal of Counseling Psychology, 1967, 14, 272-276.

Coleman, J. S., et al. Equality of educational opportunity. Washington, D. C.: U. S. Government Printing Office, 1968.

Crandall, V. C., Crandall, V. J., & Katovsky, W. A. A children's social desirability questionnaire. Journal of Consulting Psychology, 1965, 29, 27-36.

Cronbach, L. J., & Furby, L. How we should measure "change": Or should we? Psychological Bulletin, 1970, 74, 68-80.

Cox, J. B. The relationship between student ratings of teachers and value differences between teachers and students. Dissertation Abstracts, 1968, 29(5-A), 1357.

Currie, R. J. Value orientations of parents of academically successful and unsuccessful children. Dissertation Abstracts, 1967, 27(7-A), 2064.

Diener, R. E. A comparative study of selected needs, values, and attitudes of Negro and white elementary education students. Unpublished Doctoral Dissertation, University of Michigan. 1967.

Eysenck, S. B. G. Junior Eysenck personality inventory. San Diego, California: Educational and Industrial Testing Service, 1965.

Eysenck, H. J. The structure of human personality. New York: Wiley, 1960.

Flanagan, J. C. The critical incident technique. Psychological Bulletin, 1954, 51, 327-358.

Gast, D. K. Characteristics and concepts of minority Americans in contemporary children's finctional literature. Dissertation Abstracts, 1966, 27, 390-391.

Godbold, D. H. A comparison of attitudes toward school, self-perception and achievement of eighth grade students attending junior high school in communities of different levels of economic influence. Unpublished Doctoral Dissertation, University of Michigan, 1967.

Gorsuch, R. L. The general factor in the test anxiety questionnaire. Psychological Reports, 1966, 19, 308.

Gorsuch, R. L. A comparison of Biquartimin, Maxplane, Promax, and Varimax. Educational and Psychological Measurement, 1970, 30, 861-872.

Gorsuch, R. L. Educating from a contemporary theory of man as an ethical being. Educational Forum, 1972, in press.

Gorsuch, R. L., & Cattell, R. B. Second stratum personality factors defined in the questionnaire realm by the 16 P. F. Multivariate Behavioral Research, 1967, 2, 211-224.

Gross, N. English grades as a function of intellectual ability, performance and the congruency of teacher-pupil perceptions of interpersonal values. Unpublished Doctoral Dissertation, Harvard University, 1967.

Guilford, J. S., Gupta, W., & Goldberg, L. Development of a value inventory for grades 1 through 3 in five ethnic groups. Final Report, January 31, 1971, General Behavioral Systems Inc., California, Contract No. OEC-0-70-2673, U. S. Department of Health, Education and Welfare, Office of Education.

Gunderson, E. K. E., & Nelson, P. D. Life status and interpersonal values. Educational and Psychological Measurement, 1966, 34, 336-346.

Harman, H. H. Modern factor analysis. (2nd ed.) Chicago: The University of Chicago Press, 1967.

Harris, H. Development of moral attitudes in white and Negro boys. Developmental Psychology, 1970, 2, 376-383.

Heath, R. W. Ability of white teacher to relate to black students and to white students. American Educational Research Journal, 1971, 8, 1-10.

Helling, B. A. Comparative study of values among school children in Turkey and the United States. Dissertation Abstracts, 1967, 28, 497.

Hollingshead, A. B., & Redlich, F. C. Social class and mental illness: A community study. New York: Wiley, 1958.

Hummel, R., & Sprinthall, N. Underachievement related to interests, attitudes and values. Personnel and Guidance Journal, 1965, 44, 388-395.

Hunt, R. A. The interpretation of the religious scale of the Allport-Vernon-Lindzey study of values. Journal for the Scientific Study of Religion, 1968, 7, 65-78.

Hyman, H. H. The value system of different classes: A social psychology contribution to the analysis of stratification. In R. Bendix & S. Lipset (Eds.) Class, Status and Power. New York: Free Press, 1963.

Johnson, J. C., & Geoffroy, K. E. Value reflected behavioral modes: A theoretical construct. Journal of Reading Specialist, 1968, 7, 114-121.

Kagan, J., & Moss, H. A. Birth to maturity, a study in psychological development. New York: Wiley, 1962.

Keil, C. The urban blues. Chicago, Illinois: University of Chicago Press, 1966.

Kelly, F. J., & Veldman, D. T. Delinquency and school dropout behavior as a function of impulsivity and nondominant values. Journal of Abnormal and Social Psychology, 1964, 69, 190-194.

Kelly, F. J., Beggs, D. L., & McNeil, K. A. Research in the behavioral sciences: Multiple regression approach. Carbondale, Illinois: Southern Illinois University Press, 1969.

Kilpatrick, F. P., & Cantil, H. Self-Anchoring Scale, a measure of individual's unique reality worlds. Journal of Individual Psychology, 1960, 16, 158-73.

Kleinmuntz, B. Personality measurement, an introduction. Homewood, Illinois: The Dorsey Press, 1967.

Kluckhohn, F. R., & Stodtbeck, F. L. Variations in value orientations. Evanston, Illinois: Row, Peterson, 1961.

Kohlberg, L. A. Cognitive development approach to socialization. In D. A. Goslin (Ed.), Handbook of socialization theory and research. Chicago: Rand McNally, 1969.

Kohlberg, L. A. The moral atmosphere of the schools. In N. V. Overly (Ed.), The unstudied curriculum. Washington, D. C.: National Education Association, 1970.

Kohn, M. L. Class and conformity: A study of values. Homewood, Illinois: The Dorsey Press, 1969.

Landis, J. R., & Scarpitti, F. R. Perceptions regarding value orientation and legitimate opportunity: Delinquents and non-delinquents. Social Forces, 1965, 44, 83-91.

Landis, J. R., Dinitz, S., & Reckless, W. C. Differential perception of life chances: A research note. Sociological Inquiry, 1964, 34.

Larson, R. L., & Sutker, S. S. Value difference and value consensus by socioeconomic levels. Social Forces, 1966, 44, 563-569.

Liebow, E. Tally's corner: A study of Negro streetcorner men. New York: Little, 1967.

London, P. The modes and morals of psychotherapy. New York: Holt, Rinehart & Winston, 1964.

Mischel, W. Personality and assessment. New York: Wiley, 1968.

Mowrer, O. H. The new group therapy. New York: Van Nostrand, 1964.

Muncy, R. J. Value-orientations of Anglo and Navajo high school sophomores. Unpublished Doctoral Dissertation, University of New Mexico, 1967.

Nunnally, J. C. Psychometric theory. New York: McGraw Hill Book Company, 1967.

Osgood, C. E. Semantic differential technique in the comparative study of culture. American Anthropologist, 1964, 66, 171-200.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. The measurement of meaning. Urbana, Illinois: University of Illinois Press, 1957.

Peck, R. F., Havighurst, R. J., et al. The psychology of character development. New York: Wiley, 1960.

Peterson, D. R. Behavior problems of middle childhood. Journal of Consulting Psychology, 1961, 25, 205-209.

Peterson, D. R. The insecure child: Over-socialized or under-socialized? In O. H. Mowrer (Ed.), Morality and mental health. Chicago: Rand McNally, 1967.

Pettigrew, T. F. A profile of the Negro American. Princeton, N. J.: Van Nostrand, 1964.

Porter, R. B., & Cattell, R. B. Handbook for the IPAT children's personality questionnaire "The CPQ". Champaign, Illinois: Institute for Personality and Ability Testing, 1963.

Rettig, S., & Pasamanick, B. Differences in structure and severity of moral judgments by students and by a community sample: Some observations on ethical relativity. The Ohio Journal of Science, 1962, 62, 317-325.

Riessman, F. The culturally deprived child. New York: Harper & Row, 1962.

Ringness, T. A. Identification patterns, motivation, and school achievement of bright junior high school boys. Journal of Educational Psychology, 1967, 58, 93-102.

Roff, M., & Sells, S. B. Juvenile delinquency in relation to peer acceptance - rejection and socioeconomic status. Psychology in the Schools, January, 1968, 5, 3-18.

Romero, F. E. A study of Anglo - and Spanish - American culture value concepts and their significance in secondary education. Dissertation Abstracts, 1966, 27, 1559-1560.

Rokeach, M. Beliefs, attitudes, and values. San Francisco: Jossey-Bass, 1968.

Rokeach, M. Long-range experimental modification of values, attitudes, and behavior. American Psychologist, May 1971, 26, 453-459.

Rosen, B. C. Race, ethnicity and the achievement syndrome. American Sociological Review, 1959, 24, 47-60.

Rosenberg, M. Society and adolescent self-image. Princeton, N. J.: Princeton University Press, 1965.

Ryan, T. A. Value conflict in elementary schools in Hawaii. Final Report, July, 1971, Education Research and Development Center, Contract Number OEG-0-9-210528-4472, U. S. Department of Health, Education, and Welfare, Office of Education.

Savarese, J. K. The effect of the race of the experimenter on the measurement of value in children. Unpublished M. A. Thesis, George Peabody College for Teachers, 1971.

Scarpitti, F. R. Delinquent and non-delinquent perceptions of self, values and opportunity. Mental Hygiene, 1965, 49, 399-404.

Schwartz, S. H. Words, deeds, and the perception of consequences and responsibility in action situations. Journal of Personality and Social Psychology, 1968, 10, 232-242.

Scott, W. A. Values and organizations. Chicago, Illinois: Rand McNally, 1965.

Sells, S. B., Demaree, R. G., & Will, D. P., Jr. A taxonomic investigation of personality conjoint factor structure of Guilford and Cattell trait marks. Final Report, August, 1968, Texas Christian University, Contract No. OE 5-10-296, U. S. Department of Health, Education and Welfare, Office of Education.

Sherif, M., & Sherif, C. W. Reference groups: Explorations into conformity and deviation of adolescents. New York: Harper & Row, 1964.

Sherif, M., & Sherif, C. W. Social psychology. New York: Harper & Row, 1969.

Shotland, R. L., & Berger, W. G. Behavioral validation of several values from the Rokeach value scale as an index of honesty. Journal of Applied Psychology, 1970, 54, 433-435.

Smith, C. R., Williams, L., & Willis, R. H. Race, sex, and belief as determinants of friendship acceptance. Journal of Personality and Social Psychology, 1967, 5, 127-137.

Snider, J. G., & Osgood, C. E. Semantic differential technique. Chicago, Illinois: Aldine Publishing Company, 1969.

Spache, G. D. Good reading for poor readers. Illinois: The Garrard Press, 1958.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. The state-trait anxiety inventory. Palo Alto, California: Consulting Psychologists Press, 1970.

Spranger, E. Types of men. New York: Johnson Reprint Corporation, 1928.

Stromberg, R. P. Value orientation and leadership behavior of school principals. Unpublished Doctoral Dissertation, Pennsylvania State University, 1967.

Summerskill, J. Dropouts from college. In N. Sanford (Ed.), The American college. New York: Wiley, 1962.

Sundberg, N. D., Rohila, P. K., & Tyler, L. E. Values of Indian and American adolescents. Journal of Personality and Social Psychology, 1970, 16, 374-397.

Triandis, H. C., & Triandis, P. The building of nations. Psychology Today, 1969, 2, 30-35.

Welkowitz, J., Cohen, J., & Ortmeyer, D. Value system similarity investigation of patient-therapist dyads. Journal of Consulting Psychology, 1967, 31, 48-55.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

APPENDIX A

RACE OF ADMINISTRATOR EFFECT

Ruth Smith and Joan Savarese

Race of administrator was considered a possibly important variable in measuring the values of children. There is evidence in the professional literature evaluating race of experimenter effects on intelligence testing but little has been done in testing the effect of the race of the experimenter on value measurement. Therefore, several checks on the possible effects of the experimenter's race were carried out.

In Study I, race of experimenter was counter balanced across classrooms and its effects examined relative to both the race of the pupil and the race of his teacher. A multiple regression analysis from these independent variables to each value category indicated that (a) there were no significant interactions among the independent variables, and (b) E's race did not generally increase the prediction of the use of a value category over that which could be obtained from the pupil's race, his teacher's race, and sex alone. This finding, that race of administrator had no significant effect on the open-ended value eliciting questionnaire, made us question whether there would be any such effect on value measurement by an objective, close-ended, written questionnaire.

An independent project was used to test this question. It measured only academic achievement, using preliminary testing materials of the type used in Study II. Race of administrator was again counter balanced in groups containing the same number of black and white children. A total of 232 fourth, fifth, and sixth grade children were drawn from four schools in Murfreesboro, Tennessee. Approximately half of the children were males, half females. In this study the natural classroom composition of about 20% blacks and the rest whites in each classroom was changed; a sample with the same number of black and white boys and girls from the three grade levels was randomly assigned to each of the testing groups. The administrators read the materials aloud to the children.

Race of administrator did not have a significant effect on the academic achievement value scales. However, on one of two subscales, a significant interaction ($p<.05$) in an analysis of variance was found between race of administrator and sex of students. Female subjects with a black administrator scored higher than female subjects with white administrators, and higher than male subjects with a black or white administrator. The total effect was small in terms of the percent of variance influenced. As this was the only significant interaction with

race of administrator found in the two studies, and as there was no apparent explanation to it, it was concluded that race of administrator was not an important variable in measuring children's values.

APPENDIX B

Dear Parents:

Dr. Richard Gorsuch, a teacher at Peabody College, has been conducting a study in the Metropolitan School System. The purpose of the study is to find out what children see as admirable and how this influences their school life. Our school has been chosen to take part in this study.

The children will be asked some simple questions, including questions about what they like about children in general, their feelings, what they know, and what kinds of activities they like. The questionnaires will be given during school, and they should take about 1 1/2 hours at the beginning of this semester, and about 1 hour at the end of the semester. The teacher will also be asked her views of the children and what happens during classes this semester.

The study does not involve "tests" in the usual sense, as the project is not interested in "scores" on any individual child. The results will be summarized at Peabody for the children as a group, and a general report will given to the school and to you, if you wish. The school staff will not see what any child writes, nor see evaluations from these materials.

We hope that this project will eventually lead to better methods of education, and that in the future teachers will be able to make school life more pleasant for each child by better understanding children's views.

We need to know if you are willing for your child to participate. Please fill out the form below and return it to school with your child as soon as possible. I'm sure that Dr. Gorsuch will deeply appreciate your help.

Thank you.

_____ Principal

_____
Richard L. Gorsuch, Ph.D.
Associate Professor of Psychology

//

PLEASE CHECK ONE:

☐ I give permission for my child,_____,to participate in Dr. Gorsuch's study.

☐ I do not wish my child,_____, to participate in Dr. Gorsuch's study.

_____ Date

_____ Parent or Guardian

Do you wish to receive a summary report of the study? ___Yes ___ No

APPENDIX C

The Accuracy of Teachers' Ratings of Students' Social Class[1]

Richard L. Gorsuch and William L. Wallace

The John F. Kennedy Center for Research

On Education and Human Development

George Peabody College for Teachers

Lehman (1969) has shown that many techniques for estimating the socioeconomic status of children are not exceptionally accurate. For example, he found that the child's report of his father's occupation and education correlated only moderately with the same variables as determined by an interview with the father. The study concluded that the best procedure for estimating social class of children in the elementary grades was to ask them a series of questions concerning their home. Those questions included, "Do you have a telephone?" and "Do you have more than thirty books in your house?" Even with a series of these questions, the multiple correlation with social class determined by home interview was only about .6.

Classically, social status has been defined by how others in the community view the individual. If a person is highly respected and valued by others, then he is higher in social class than those who are less respected. Measures of socioeconomic status (SES) such as education, occupation and income have become widely used because studies such as that of Hollingshead and Redlich (1956) have found them to be closely identified with the community opinion.

But since it is difficult to gather accurate information about their parents' status from the children themselves, it might be better to go to a member of the community who has some knowledge of the children and their situations. The child's teacher is one such person.

If teachers can estimate the children's social class with reasonable accuracy, then such a resource might be preferred to interviews with either the parents or the child. Interviewing parents is often costly, and asking the child questions about his home may be considered an invasion of privacy. Having the teacher estimate the child's social class appears to be relatively inexpensive and reduces the invasion of privacy problem.

It is scarcely expected that teachers' perceptions of class will be 100% accurate. Interpersonal judgments are always confounded by a host of irrelevant variables. To the extent that it is possible to identify such determinants of teachers' estimates of social class, corrective procedures can be employed.

Investigating the accuracy of teachers' social class ratings is also important since there is an increasing body of literature which suggests that a teacher's expectations of the child's behavior will influence her behavior toward the child (e.g., Dalton, 1968; Pitt, 1965; Rosenthal & Jacobson, 1968a, 1968b, 1966). If teachers' expectations for children differ according to social class, as Lenkowsky and Blackman (1968) have found, it is important to know the accuracy of their perception of this variable. If their perceptions are inaccurate, then their expectations will probably be inappropriate.

Among the factors which may affect the teacher's accuracy is the use of racial stereotypes. For example, it would be important if it were found that the teacher systematically misjudges the SES of black children, and such an effect would be compounded if white teachers' errors differed from those of black teachers.

The percentage of black pupils in a classroom could also affect the teacher's estimate of social class. If there are only one or two

black children, the teacher might tend to judge them as individuals and, therefore, offer a more accurate appraisal of their social class. If, however, there is a larger number of black children--yet still a distinct minority--the teacher might judge them as a group and rate them lower than comparable white children. When the percentage of black children becomes a distinct majority, then one might expect the effect of comparison with the whites to become minimal again. Given a majority of blacks in a classroom, one would expect them to become the norm for class.

The present study was undertaken to answer two basic questions concerning teacher's estimates of social class: (1) How accurate are teachers' estimates? (2) To what extent are the estimates distorted by race?

Subjects and Method

The sample was 344 fifth and sixth grade children from three public schools in a Southern city. Five teachers in school 1, five in school 2, and three teachers in school 3 rated their children by placing them in one of three categories: "middle class," "lower class," or "don't know." The categories were not further defined. These categories were deemed sufficiently accurate since the sample was known to contain few upper class students and since the raters were not given the training necessary for finer discriminations. (In another study, finer discriminations with technical categories were attempted but the results were disappointing.) The teacher also indicated the child's race.

The name of the head of the household was obtained from the school records for 300 of the original 344 Ss. Of the 44Ss who were dropped, eight were omitted because of miscellaneous reasons, such as missing data. Another 36 were dropped from the sample because their parents'

names were not available in the school records.

Each parent's name was submitted to a credit bureau, which rated SES on a 7-point scale on the basis of the father's occupation. The 7-point occupational scale for socioeconomic status was that developed by Hollingshead and Redlich (1958).

Where needed, the occupational scale was collapsed into two categories in order to be comparable with the teachers' categories of middle and lower class. Hollingshead's occupational categories 1, 2, 3, and 4 were grouped together as middle class since these represent the general description of "white collar" occupations (e.g., professionals, owners and managers of businesses, clerks and technicians). Categories 5, 6, and 7 were grouped together as characterizing lower class or "blue collar" workers (e.g., skilled, semi-skilled, and unskilled manual employees).

For most analyses, the subject pool was reduced by eliminating subjects rated "don't know" by the teacher or for whom occupation information could not be determined. This left 214 Ss in the subject pool for detailed analysis. The sample included 62 black males, 68 black females, 41 white males, and 43 white females.

Analysis and Results

To determine whether the occupational categories of "don't know" and "know" were biased toward a particular social class, Table 1 was constructed. It shows the relationship between the 336 teachers' ratings and the occupational ratings. The supposition was that it would be more difficult to find employment information for lower class fathers. The results indicate the contrary: the distribution of occupational "don't knows" across the three categories of the teachers' ratings was not significantly different from the distribution of known occupational categories.

Table 1 goes about here

A multiple regression analysis (Kelly, et al., 1969; Mendenhall, 1968) was run to assess the extent to which the percentage of black children in a classroom, the race of the pupil, the race of the teacher, and the teacher's estimate of social class were predictors of the socioeconomic status. Multiple regression analysis is an elaboration of the linear model, $Y = a + bX$, which makes it possible to combine several variables and use them collectively as predictors of a criterion variable such as occupational SES. The model can include terms which could account for possible interactions and curvilinear effects. By systematically removing variables from the full model, it becomes possible to judge the amount of information each variable adds to the estimate of SES over and above the set of other variables.

Table 2 summarizes the multiple regression analyses. The first column gives the source of the effect; the second gives the additional percent of variance that could be predicted by adding that variable to the model, and the last column gives the significance of that effect. All the rows taken together represent the full model against which the reduced models were tested.

Table 2 goes about here

The strongest relationship found was between the teacher's estimate of social class and the occupational SES. While this relationship was not huge, it was highly significant ($p < .001$) and large enough to be useful for research purposes. It was equivalent to a biserial correlation of .625.

Table 2 indicates that the effect of percent of black pupils in

the classroom and the interaction of pupil's and teacher's race accounted for a small and non-significant part of the total variance. Race of teacher and race of pupil contributed even less in terms of main effects. Since none of these possible moderator variables was significant, the remaining statistical analyses ignored race of pupil, race of teacher, and the percent of blacks in the classroom.

The accuracy of teachers' estimates of class was then examined more closely. Contingency tables and chi-squares were calculated between the teacher's estimates of social class and the occupational estimate of SES for each school. These are presented in Table 3.

The significant relationship between teachers' ratings and SES was found in schools 1 and 2, but not in school 3. Such a lack of relationship appears to be a function of the high homogeneity in school 3. One suspects that the middle class children were lower middle class at best and consequently difficult to distinguish. Note that even in that case the accuracy was high (80%), and the accuracy rate was actually higher than that found in the other two schools ($X^2 = 7.34$, df = 2, $p < .05$). School 1 had the lowest accuracy; it was the largest and most heterogeneous.

In total, the teachers' estimates correctly classified 75% of the children whom they rated. The reputations of the schools lead to only 61% of the Ss being correctly classified.

Table 3 goes about here

Discussion

Teachers can often estimate the social class of their pupils. While the estimates in this study were not perfect, they did indicate that 75% of the children rated were correctly identified as lower or

middle class. The teachers' estimates therefore compare favorably with other measures suggested for this age child. Lehmann (1969) found that directly asking eighth grade children about their parents' SES status gave accuracy rates up to those found in this study, but not higher; Lehmann's data suggests that asking fifth and sixth grade children such questions would be less effective. Asking about characteristics of the home (e.g., "Does your family have an encyclopedia?") leads to a multiple correlation of .63 in that same study, a coefficient which is quite similar to our biserial correlation of .625. In addition, Lehmann found some of the questions asked children were differentially effective as a function of the child's social class, a problem not encountered with teachers' ratings. Since the ratings of the present study occurred early in the school year (October), accuracy and confidence could increase as the teacher becomes more familiar with her students.

The teachers did contribute more information for less cost than did the occupational method. Of the original sample, 24% of the subjects were dropped because of insufficient occupational information while only 18% of the original sample of 344 had to be dropped because the teacher did not know the pupil's social class. These percentages include some overlap; about 5% of the total Ss could not be placed in a social class by either the occupational method or by the teacher.

When research needs require a simple yet relatively effective method of obtaining the SES of children, teachers' estimates appear to offer several advantages: (1) Teachers' ratings of social class show an accuracy rate which appears to be about the same as the best methods known to the writers that do not interview the parents. (2) Variables such as the teacher's race and the pupil's race do not appear to be

distorting teachers' estimates of class, even in a Southern city. (3) The method avoids the invasion of privacy problems involved when children are interrogated with respect to their homes. (4) Its costs are small compared to the cost of employing a credit bureau to ascertain head-of-household's occupation. Also, (5) fewer subjects need to be dropped due to insufficient information than is the case when occupational ratings of SES are used.

References

Dalton, W. B. The relations between classroom interaction and teacher ratings of pupils: An exploration of one means by which a teacher may communicate her expectancies. Unpublished Master Thesis, George Peabody College for Teachers, 1968.

Hollingshead, A. B. & Redlich, Frederick C. Social class and mental illness, New York: Wiley, 1958, pp. 387-397.

Kelly, F. J., Beggs, D. L., McNeil, K. A., et al. Research design in the behavioral sciences: Multiple regression approach. Carbondale, Illinois: Southern Illinois University Press, 1969.

Lehmann, I. Correlates of socio-educational stutus. Cape Pilot Study, Research Triangle Institute. July, 1969.

Lenkowsky, Leonard, & Backman, L. S. The effect of teachers' knowledge of race and social class on their judgments of children's academic competence and social acceptability. Mental Retardation, 1968, 6, 15-17.

Mendenhall, W. Introduction to linear models and the design and analysis of experiments. Belmont: California, Wadsworth Publishing Company, Inc., 1968.

Pitt, C. C. V. An experimental study of the effects of teachers' knowledge or incorrect knowledge of pupil I.Q.'s on teachers' attitudes and practices and pupils' attitudes and achievement. Unpublished doctoral dissertation, Columbia University, 1956.

Rosenthal, R. & Jacobson, Lenore. Teacher's expectancies as determinants of pupils' I.Q. gains. Psychological Reports, 1966, 19, 115-118.

Rosenthal, R. & Jacobson, Lenore. Pygmalion in the classroom. New York: Holt, Rinehart & Winston, 1968.

Rosenthal, R. & Jacobson, Lenore. Self-fulfilling prophecies in the classroom: Teachers' expectations as unintended determinants of pupils' intellectual competence. In M. Deutsch, I. Katz, & A. Jensen (Eds.), Race, social class, and psychological development. New York: Holt, Rinehart & Winston, in press.

Footnotes

[1]The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

Table 1

Distribution of Teacher's Estimates of Social Class

for Subjects Whose Occupational SES was Unknown

Teacher Rating

(N = 336)

Occupational Rating	Don't know	Lower	Middle	
Don't know	5%	13%	6%	24%
Know	13%	39%	24%	76%
	18%	52%	30%	

$X^2 = 4.07$, df = 2, $p > .1$

Table 2

Summary of Multiple Regression Analysis for the

Prediction of Occupational SES

Source of effect (Predictors)	Percent of variance accounted for	Significance
Teachers' estimate of SES	24.4%	p<.001
Percent of black pupils in the classroom	.1%	NS
Race of pupil	.0%	NS
Race of teacher	.4%	NS
Quadratic effect of % black, and teacher's race X pupil's race interaction	.5%	NS
Total variance accounted for	25.4%	p<.0001

Table 3

Relationship of Teacher's Ratings of Pupil's

Social Class and Occupation SES

A. School 1 (N = 78)

Occupational Ratings	Teacher Ratings: Lower	Teacher Ratings: Middle	
Middle	6%	35%	41%
Lower	30%	30%	60%
	36%	65%	101%

Teacher's Accuracy Rate: 65%

$X^2 = 9.69$; $p < .005$

B. School 2 (N = 86)

Occupational Ratings	Teacher Ratings: Lower	Teacher Ratings: Middle	
Middle	8%	21%	29%
Lower	59%	12%	71%
	67%	33%	100%

Teacher's Accuracy Rate: 80%

$X^2 = 24.97$; $p < .001$

C. School 3 (N = 50)

Occupational Ratings	Teacher Ratings: Lower	Teacher Ratings: Middle	
Middle	10%	0%	10%
Lower	82%	8%	90%
	92%	8%	100%

Teacher's Accuracy Rate: 82%

$X^2 = .48$; $p < .50$

D. All Schools (N = 214)

		Teacher Ratings: Lower	Teacher Ratings: Middle	
Occupational Ratings	Middle	8%	21%	29%
	Lower	54%	17%	71%
		62%	38%	100%

Teacher's Accuracy Rate: 75%

$X^2 = 43.36$; $p < .001$

APPENDIX D

VALUE MEASUREMENTS DEVELOPED BY THE PROJECT

1. Value Socialization Scale

Soc.9/70

NAME_____

In this part you are asked to circle one answer that fits how you feel about what a child may be or what he may do. The answers from which you will choose are the following:

I always admire, I sometimes admire, I always dislike

"Admire" means to think someone is special, to respect someone for doing something, and to want to be like that person.

Example:

A CHILD'S EATING EVERYTHING FROM HIS PLATE.

1. I always admire 2. I sometimes admire 3. I always dislike

Circle only one answer for each question. Remember, this is not a test. There is no right or wrong answer. Some children admire a child who eats everything from his plate, and some dislike such a child. Both are fine, for that's the way they feel. Your answer will be fine if it tells how you really feel.

LET US START WITH THE FIRST QUESTION.

1. A CHILD'S HELPING EVERYBODY WHO NEEDS HELP.

1. I always admire 2. I sometimes admire 3. I always dislike

2. A CHILD'S DOING WHAT HE WANTS INSTEAD OF WHAT HIS PARENTS ASK.

1. I always admire 2. I sometimes admire 3. I always dislike

3. A CHILD'S DOING THE BEST HE CAN.

1. Always admire 2. Sometimes admire 3. Always dislike

4. A CHILD'S GETTING BY IN SCHOOL WITH LITTLE WORK.

1. Always admire 2. Sometimes admire 3. Always dislike

5. A CHILD'S GIVING WRONG ANSWERS ON A TEST.

1. Always admire 2. Sometimes admire 3. Always dislike

6. A CHILD'S DOING WHAT THE TEACHER WANTS.

1. Always admire 2. Sometimes admire 3. Always dislike

7. A CHILD'S TALKING BACK TO HIS PARENTS.

1. Always admire 2. Sometimes admire 3. Always dislike

8. A CHILD'S BEING KIND TO OTHERS.

1. Always admire 2. Sometimes admire 3. Always dislike

9. A CHILD'S TALKING BACK TO THE TEACHER.

1. Always admire 2. Sometimes admire 3. Always dislike

10. A CHILD'S GOING TO THE STORE FOR SOMEBODY.

1. Always admire 2. Sometimes admire 3. Always dislike

APPENDIX D
2. Semantic Differential Value Scale

SD/F 70

MEANINGS

NAME _____ GRADE _____
First Middle Last

BOY OR GIRL _____ AGE _____

In this questionnaire you are asked to think about what some children are or what they do in a new and different way. You are to use your imagination and tell what certain words mean to you.

Example: The words I tell you are:

Getting new shoes

Then you will find several answers on a line

1.very happy 2.happy 3.so-so 4.sad 5.very sad

Circle the one answer from each line that best describes getting new shoes.

Now is getting new shoes:

1.very bad 2.bad 3.so-so 4.good 5.very good

Again, circle the answer that best fits getting new shoes.

Then I'll ask you about something else, for example:

Telling a lie

Again, please circle one answer on each line.

1.very small 2.small 3.so-so 4.big 5.very big

1.very strong 2.strong 3.so-so 4.weak 5.very weak

Now, let's start with the first question.

PLEASE ANSWER EACH QUESTION BY CIRCLING ONLY ONE ANSWER ON EACH LINE.

DOING WELL IN SCHOOL WORK (for example: studying hard to become one of the better students in class, finishing your homework even if it means missing a TV show, liking to do extra work for school).

I.	1.very dull	2.dull	3.so-so	4.sharp	5.very sharp
II.	1.very good	2.good	3.so-so	4.bad	5.very bad
III.	1.very sad	2.sad	3.so-so	4.happy	5.very happy
IV.	1.very awful	2.awful	3.so-so	4.nice	5.very nice
V.	1.very small	2.small	3.so-so	4.big	5.very big
VI.	1.very strong	2.strong	3.so-so	4.weak	5.very weak

BEING WELL-LIKED BY OTHERS (for example: sharing special secrets, having a lot of friends, being liked by other children).

I.	1.very strong	2.strong	3.so-so	4.weak	5.very weak
II.	1.very dull	2.dull	3.so-so	4.sharp	5.very sharp
III.	1.very awful	2.awful	3.so-so	4.nice	5.very nice
IV.	1.very sad	2.sad	3.so-so	4.happy	5.very happy
V.	1.very good	2.good	3.so-so	4.bad	5.very bad
VI.	1.very small	2.small	3.so-so	4.big	5.very big

OBEYING PEOPLE WHO KNOW BETTER (for example: doing what the teacher tells you, obeying school rules, stopping play to do what you are told).

I.	1.very dull	2.dull	3.so-so	4.sharp	5.very sharp
II.	1.very good	2.good	3.so-so	4.bad	5.very bad
III.	1.very sad	2.sad	3.so-so	4.happy	5.very happy
IV.	1.very strong	2.strong	3.so-so	4.weak	5.very weak
V.	1.very awful	2.awful	3.so-so	4.nice	5.very nice
VI.	1.very small	2.small	3.so-so	4.big	5.very big

HAVING GOOD MANNERS (for example: saying "good morning" to the teacher, getting out of the classroom quietly when dismissed, waiting in line for your turn).

I.	1.very small	2.small	3.so-so	4.big	5.very big
II.	1.very strong	2.strong	3.so-so	4.weak	5.very weak
III.	1.very sad	2.sad	3.so-so	4.happy	5.very happy
IV.	1.very good	2.good	3.so-so	4.bad	5.very bad
V.	1.very dull	2.dull	3.so-so	4.sharp	5.very sharp
VI.	1.very awful	2.awful	3.so-so	4.nice	5.very nice

BEING HONEST (for example: bringing back correct change from the store, being able to admit you did something wrong, keeping your promise).

I.	1.very good	2.good	3.so-so	4.bad	5.very bad
II.	1.very small	2.small	3.so-so	4.big	5.very big
III.	1.very strong	2.strong	3.so-so	4.weak	5.very weak
IV.	1.very sad	2.sad	3.so-so	4.happy	5.very happy
V.	1.very awful	2.awful	3.so-so	4.nice	5.very nice
VI.	1.very dull	2.dull	3.so-so	4.sharp	5.very sharp

BEING KIND TO OTHERS (for example: thinking of others' feelings, breaking up fights, making others happy).

I.	1.very sad	2.sad	3.so-so	4.happy	5.very happy
II.	1.very small	2.small	3.so-so	4.big	5.very big
III.	1.very strong	2.strong	3.so-so	4.weak	5.very weak
IV.	1.very good	2.good	3.so-so	4.bad	5.very bad
V.	1.very awful	2.awful	3.so-so	4.nice	5.very nice
VI.	1.very dull	2.dull	3.so-so	4.sharp	5.very sharp

BEING GOOD LOOKING (for example: being clean, spending time taking care of the way you look, getting a good grade in Physical education).

I.	1.very good	2.good	3.so-so	4.bad	5.very bad
II.	1.very sad	2.sad	3.so-so	4.happy	5.very happy
III.	1.very small	2.small	3.so-so	4.big	5.very big
IV.	1.very dull	2.dull	3.so-so	4.sharp	5.very sharp
V.	1.very awful	2.awful	3.so-so	4.nice	5.very nice
VI.	1.very strong	2.strong	3.so-so	4.weak	5.very weak

TRYING TO ALWAYS DO YOUR BEST (for example: finishing every job you start, giving up playing when you have a task to do, working on your job even when it is dull).

I.	1.very strong	2.strong	3.so-so	4.weak	5.very weak
II.	1.very sad	2.sad	3.so-so	4.happy	5.very happy
III.	1.very dull	2.dull	3.so-so	4.sharp	5.very sharp
IV.	1.very awful	2.awful	3.so-so	4.nice	5.very nice
V.	1.very small	2.small	3.so-so	4.big	5.very big
VI.	1.very good	2.good	3.so-so	4.bad	5.very bad

All materials included here have been prepared by Richard L. Gorsuch, Peabody College. Do not use without permission.

APPENDIX D

3. Rank Ordering of Eight Values
Instructions

(After the statements on the following page are read, the instructor gives the following instructions verbally).

Which one of these do you most admire? You may, of course, admire them all, but we want to know which one you like more than the others. Choose the one you admire the most and print a 1 on the line in front of it.

Which one would be your second choice? Print a 2 on the line in front of that one.

Now look at the rest of the statements. Which one do you like most of those which you have not already chosen? Print a 3 in front of that one.

Now print a 4 in front of the one you like next.

Pick the one you like next and put a 5 on the line in front of it.

Write a 6 in front of the one you next admire.

There are two remaining statements. Write a 7 on the line before the one you admire more than the other.

Finally print an 8 on this line before the last statement. This should be the statement you like the least.

NAME _____ Rank Order 9/70

On this page there are 8 things that a child may do that most children admire. Let's first read all of them.

_____ A CHILD'S BEING KIND TO OTHERS (for example, thinking of others' feelings, breaking up fights, making others happy).

_____ A CHILD'S OBEYING PEOPLE WHO KNOW BETTER (for example, doing what the teacher tells him, obeying school rules, stopping play to do what he is told).

_____ A CHILD'S BEING HONEST (for example, bringing back the correct change from the store, being able to admit he did something wrong, keeping his promise).

_____ A CHILD'S BEING WELL LIKED BY OTHERS (for example, sharing special secrets, having a lot of friends, being liked by other children).

_____ A CHILD'S BEING GOOD LOOKING (for example, being clean, spending time taking care of the way he looks, getting a good grade in Physical Education).

_____ A CHILD'S DOING WELL IN SCHOOL WORK (for example, studying hard to become one of the better students in class, finishing his homework even if it means missing a TV show, liking to do extra work for school).

_____ A CHILD'S TRYING TO ALWAYS DO HIS BEST (for example, finishing every job he starts, giving up playing when he has a task to do, working on his job even when it is dull).

_____ A CHILD'S HAVING GOOD MANNERS (for example, saying "good morning" to the teacher, getting out of the classroom quietly when dismissed, waiting in line for his turn).

APPENDIX E

SPECIAL MEASURING INSTRUMENTS USED BY THE PROJECT

1. Perceived Potency of Teacher*

MY TEACHER

I.	1.very strong	2.strong	3.so-so	4.weak	5.very weak
II.	1.very dull	2.dull	3.so-so	4.sharp	5.very sharp
III.	1.very rough	2.rough	3.so-so	4.smooth	5.very smooth
IV.	1.very hard	2.hard	3.so-so	4.soft	5.very soft

* Was given as part of the Semantic Differential test.

APPENDIX E

2. Child's Peer Group

W/71

NAME _____

PEER NOMINATION

1. Please list the first names of children with whom you go around. That is, the children you spend time with after school, the ones you do things with and like to be with.

2. Are any of the children you listed above in your classroom? If yes add their last names to the list.

3. Please list the children in this classroom with whom you would most like to go around. Again, please write both their first and last names.

4. Please list the children in the classroom with whom you've gone around since school started in September. Again, please write both their first and last names.

APPENDIX E

3. Self-Anchoring Scale F/70

Instructions for the Self-Anchoring Scale

A self-anchoring scale is simply one in which you yourself define the extremes, or "anchors," of a dimension and then use this self-defined continuum as a rating scale.

In the space provided below first describe what, for you, would be the ideal student, or the student whom you would most admire. Take care to record as many examples as you can.

Now describe what you would consider to be the student whom you would least admire. Again, think of as many examples as you can.

You are now ready to turn the page and rate your pupils on the nine-point ladder scale provided. Your description of the ideal or most admired student represents a rating of 9, and the description of the least admired student, a rating of 1.

For each pupil, place a check under his name in the row corresponding to the rating you think he should receive.

Ideal or most admired student:
Least admired student:

Self-Anchoring Scale

RATING	EXAMPLES
9	Ideal or most admired student
8	
7	
6	
5	
4	
3	
2	
1	Least admired student.

APPENDIX E

4. Value Ratings

F/70

Instruction for the Value Rating Scale

In the following scales you are asked to rate each of your pupils on eight value dimensions. Example definitions are given for the "extremes," or 1 and 9 ratings.

For each pupil, place a check under his name in the row corresponding to the rating you think he should receive.

PLACE NAME LIST HERE

Value Rating Scale

Academic Achievement

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	RATING	EXAMPLES
																														9	A child who works hard to understand difficult lessons and books and who trys to maintain good grades.
																														8	
																														7	
																														6	
																														5	
																														4	
																														3	
																														2	
																														1	A child who seems to dislike studying, does not care what grades he makes, and is proud to get by in school with little work.

88+

PLACE NAME LIST HERE

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

RATING — Value-Rating Scale

Affiliation

EXAMPLES

9 — A child who seems to enjoy being with others and who has many friends or belongs to a "special" group.

8

7

6

5

4

3

2

1 — A child who seems to need few friends and who enjoys playing alone.

PLACE NAME LIST HERE

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

190

RATING	EXAMPLES
9	A child who finishes every job and who works to get a job done even when it is dull.
8	
7	
6	
5	
4	
3	
2	
1	A child who often does a sloppy job and plays when he should be working.

Value-Rating Scale

Diligence

Value-Rating Scale

Good Manners

PLACE NAME LIST HERE

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

101

RATING	EXAMPLES
9	A child who is always polite and courteous, showing good manners in all his interactions with others.
8	
7	
6	
5	
4	
3	
2	
1	A child who is mostly impolite and rude in his dealings with others.

PLACE NAME LIST HERE

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

RATING	Value-Rating Scale Honesty EXAMPLES
9	A child who is always honest and is able to admit that he did something wrong.
8	
7	
6	
5	
4	
3	
2	
1	A child who often breaks promises and lies to the teacher about what he has been doing.

PLACE NAME LIST HERE

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Value-Rating Scale

Kindness

RATING	EXAMPLES
9	A child who thinks of others' feelings and often helps them and makes them happy.
8	
7	
6	
5	
4	
3	
2	
1	A child who hurts others' feelings, talks behind their backs, and tells stories about them.

1933

PLACE NAME LIST HERE

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Fig. 4

RATING	Value-Rating Scale Obedience EXAMPLES
9	A child who always follows directions, obeys school rules, and does what he is told to do.
8	
7	
6	
5	
4	
3	
2	
1	A child who often does things the teacher has said not to do and does what he wants instead of listening to people who know better.

PLACE NAME LIST HERE

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	193																												

Value-Rating Scale

Physical Appearance and Development

RATING	EXAMPLES
9	A child who is clean, pleasant to look at, and who spends time taking care of the way he looks. A child who is good at sports.
8	
7	
6	
5	
4	
3	
2	
1	A child who wears messy clothes, has a dirty appearance and often gets a poor grade in physical education.

APPENDIX E

5. Teacher's Estimate of SES

Problem Check List

Student's Name:_____Grade:_____Race:_____

Social Class: Middle_____ Lower_____ Don't Know_____

Teacher doing the rating_____

Below are thirty phrases describing characteristics of children. Place a 1 before each term that is a problem with the child. If the term or phrase is not descriptive of the child, place a 0 in front of it. Please note that for some children there will be only 0's on the list.

_____feels inferior

_____lacks self confidence

_____withdraws socially

_____prone to become frustrated

_____self conscious

_____shy

_____anxious

_____lethargic

_____unable to have fun

_____depressed

_____reticent

_____hypersensitive

_____drowsy

_____aloof

_____preoccupied

_____disobedient

_____disruptive

_____boisterous

_____fights

_____seeks attention

_____restless

_____negative

_____impertinent

_____destructive

_____irritable

_____has temper tantrums

_____hyperactive

_____uses profanity

_____jealous

_____uncooperative

APPENDIX E

6. Important Incident Report F/70

Instructions for the Recording of
Important Classroom Incidents

Important classroom incidents are those which, as you look back over the day, stand out. They are the incidents that made you feel either particularly "pleased" or "disturbed" during the course of a school day.

The number of such incidents per day may vary from none at all to several.

These important incidents should be recorded on the forms provided at the end of each school day. As a guide, you may find it helpful to follow the procedure outlined below.

(1) Sit down and relax. Think back over the day.

(2) As the outstanding pleasing or disturbing incidents come to mind, write down the name of the pupil involved.

(3) Using the list of names as a reminder, fill out the Important Incident forms.

E-17

Date_____

Time_____

Important Incident

Pupil_____Teacher_____

Class Activity_____

What happened?

What did you do?

What was the other pupils' reaction, if any?

APPENDIX F

DANGERG IN USING THE CHILDREN'S LOCUS OF CONTROL SCALE WITH CHILDREN

Richard L. Gorsuch, Richard P. Henighan, Caroline Barnard

George Peabody College for Teachers

Beginning with Rotter's discussion of social learning theory (1954), investigators have stressed that people vary in their expectancies for reinforcement. At one extreme are those who feel that the reinforcements which they receive are completely within their power; they perceive the locus of control of that reinforcement as internal. At the other extreme are those who feel that their reinforcements are controlled by factors which they cannot influence; they perceive the locus of control of their reinforcements as external to themselves. As the reviews by Rotter (1966) and Lefcourt (1966) point out, distinguishing internal and external loci of control has been useful for many research purposes. Rotter (1966) presents information on an adult locus of control scale which suggests that the concept can be readily operationalized.

Children's versions of locus of control scales have been prepared. Bialer (1961) adapted items from studies of older youth (e.g., James , 1957) for this purpose. He reported that individually administering this 23 item scale to educable mentally retarded and normal children resulted in an adequate split-half reliability (Bialer, 1960). Miller (1960) also found an acceptable level of reliability in a sample of mentally retarded teenagers, whose mental ages averaged about 9 to 10 years. The Spearman-Brown corrected split-half reliability was .87. From evidence such as this, Cromwell, et al, (1961) suggested that this scale can be used with subjects who have MA's as low as 4 years.

A 20 item children's locus of control scale was developed by Miller (1963). He found an internal consistency reliability of .89 for 98 fifth

graders on this revised version of Bialer's scale. Butterfield (1963) also found the reliability of this revised scale to be adequate for use with children (r_{xx} = .84). Blanton and Nunnally (1964) state that the scale is usable with children in the fourth grade.

The reliabilities reported above were calculated on total samples of subjects. This procedure can be dangerous since a subgroup of the children may not have as high a reliability as the rest. This would be the case if the verbal content were too difficult for some of the children. Since low verbal ability children could not understand some of the items, they would respond in a more random fashion and so have less reliable scores. Interpreting data in such a situation would be appropriate only if it could be assumed that the relationship of locus of control to other variables is not greatly affected.

The possibility that a scale may require greater verbal skills than some of the children have can be checked. To do so, the children are divided into several groups according to their verbal ability. Within each of the homogeneous ability groups, the reliability of the scale is calculated. If the scale is too difficult for some ability levels, then those groups will have lower reliabilities.

For illustrative purposes, the results of such analyses on two scales designed for children are presented in Figure 1. The total pool of fourth and fifth grade children was divided into 10 ability groups (one was low) on the basis of their scores on a verbal analogies test. Reliabilities of the children's social desirability scale (Crandall, Crandall, & Katkovsky, 1965) and a socialization scale developed by Gorsuch and Smith (personal communication) were calculated within each of the verbal ability subgroups. As can be seen in Figure 1, the socialization scale reliabilities form a flat straight line with only random departures

from that line. This indicates that all of the ability groups were able to comprehend it equally well. However, the best fitting "curve" for the social desirability scale is a straight line which has a slope (or correlation) significantly different from the horizontal line one would find if the reliabilities had only chance departures from the average. The resulting correlation of .71 between ability group and reliability indicates that the lower ability groups have poorer reliabilities than the better ability groups. Fortunately, even the poorest groups have good reliabilities, which indicates that most of the social desirability scale is not too difficult. There appears to be no real danger involved in using these two scales with children.

insert Figure 1 about here

In both of these examples, the relationship between ability and reliability is best described by a straight line. This may not always be the case. If the scale has items of uniform verbal difficulty, the low ability groups could have low reliabilities while the moderate and high ability groups may all understand the items and have the same high reliabilities. Plotting these reliabilities as a function of ability groups would produce a curve. This curve would be described by a quadratic equation rather than the usual linear one. So both linear and quadratic equations need to be tested for fit to the data when testing for low reliabilities from a lack of verbal ability. In both of the illustrative cases, the quadratic equation did not significantly reduce the error of fit over the linear equation.

The present studies computed reliabilities within each of several ability groups to check the adequacy of a children's locus of control scale. The first study used the Bialer scale and the second used Miller's

revision of that scale.

Study One

To examine the reliability characteristics of the Bialer locus of control scale for children, it was administered twice to elementary school children. These children were from one fourth and one fifth grade classroom from each of eight schools. The schools were selected such that two were from each of the following categories: lower class black, lower class white, middle-class black and middle-class white. The sample, therefore, probably included more low verbal ability children than would be found in a random sample from these grades. If the scale functioned adequately for this sample, it could probably be assumed to function as well with samples drawn on a more random basis.

In the first administration, Bialer's locus of control items were mixed with those from the social desirability scale and Eysenck's introversion scale for children (Eysenck, 1965). The scales were read aloud to each classroom of children. Half of the items of this omnibut inventory were given before thirty minutes of additional testing materials and half were given after the additional materials. A counterbalanced order was used with each half being given first in 8 of the 16 classrooms. The additional materials included scale B of the Children's Personality Questionnaire (Potter & Cattell, 1963), which is a measure of verbal ability.

The children were split into 10 ability groups on the basis of their scores on scale B. (The number of subjects in each of the ten groups is given at the bottom of Figure 2.) An internal consistency reliability, coefficient alpha (Nunnally, 1967), was calculated within each ability group for the locus of control scale. The reliabilities for the first

administration of the scale were plotted in Figure 2 as a function of ability. The best fitting curve was determined by multiple regression procedures. A quadratic curve fit the data above chance level ($R = .76$, $p < .05$) and gave a significantly better fit than only a straight line (F of difference = 8.69, df = 1/7, p = .02).

Insert Figure 2 about here

Obviously, when the Bialer locus of control scale is mixed with other scales and given to fourth and fifth graders, the reliability is very poor overall and is even worse with some ability groups. Could this have occurred because the scale was mixed with other scales? This possibility exists even though the other scales had appropriate reliabilities in this situation (e.g., the social desirability scale from the same omnibus inventory was plotted in Figure 1 above).

To check the effects of mixing the items, the scale was readministered to the same children after a three month interval - this time it was give alone. The number of subjects in each ability group dropped slightly due to normal attrition; these N's are also given at the bottom of Figure 2.

The results of calculating the reliabilities for the second administration were plotted in Figure 2. The correlation of each group's ability score with its reliability was .95 ($p < .001$); a quadratic equation did not significantly improve the fit over that linear equation (F of the difference = 1.22, df = 1/7, $p > .3$). Again the scale was obviously not appropriate for the lower ability groups and was of only limited value for the brighter fourth and fifth graders. The scale did appear to function somewhat more reasonably when it was not mixed in an omnibus inventory but this improvement is not such as to greatly encourage use of the scales.

The test-retest reliabilities are not presented. They fell between the two sets of internal consistency reliabilities and led to the same conclusions.

Study Two

The second study used Miller's revision of the Bialer locus of control scale for children. The subjects were sampled from 17 fourth grade classrooms in eight schools stratified according to class and race; only boys were used in this study. The test was administered in the same session as other research instruments; its items were mixed only with locus of evaluation items. The materials were read aloud to the children.

The subjects were grouped on the basis of their language score from the California Test of Mental Maturity. The number of subjects in each of the ability categories is given at the bottom of Figure 3.

Insert Figure 3 about here

The results of the second study are summarized in Figure 3. Curve fitting procedures indicated that the ability level of the groups correlated .79 ($p < .01$) with their reliabilities, but that a quadratic equation did not significantly increase the accuracy of the fit ($F = 0.09$ $df = 1/5$). Note that the reliability was non-existent for the lowers ability group and poor for even those groups with higher verbal skills.

Discussion

The results show that the Bialer and Miller children's locus of control scales are not appropriate for all fourth and fifth grade children. In each of the three administrations, the locus of control scales' internal consistency reliabilities ranged from non-existent to poor for the low ability fourth and fifth graders. For the high ability groups,

the reliabilities ranged from poor to acceptable for research purposes. This means that those children with poorer reading skills are likely to respond randomly even when the items are read aloud to them. Such a scale should be reserved for use with older children and teenagers or for use in situations where it can be shown that the children have better verbal skills than is usual for their age. Even in situations where the latter seems to be true, it would be best to evaluate the reliability of the scale to be sure it is functioning appropriately.

There are subtle dangers involved in using any scale where the reliability varies as a function of another characteristic. In that situation, relationships with other variables will be a function of both that which is measured by the scale and the moderating characteristic. Correlations and ANOVA's would be ambiguous since any relationship found could be a function of the interaction between the concept being measured and the moderator, even though the relationship would usually be interpreted as solely attributable to the concept.

For example, assume that the children's locus of control scale is used in a situation where those low in verbal ability had totally unreliable scores and those high in verbal ability had quite reliable scores. If the subjects are divided at the median on locus of control, the high group would contain those with an internal locus of control but who also have high verbal ability. The low group will contain those with an external locus of control and high verbal ability. Those low in verbal ability will be randomly placed in the internal and external locus of control groups since their locus of control scores are unreliable. Any significant difference found between the two groups would be a difference between only those high in verbal skills. Comparing these two groups for

differences on another variable will be legitimate only if it is reasonable to assume that the other variable relates to locus of control in the same manner at low levels of verbal ability as it does at high levels. It would be difficult to test this assumption or to control statistically for the impact of ability if the assumption cannot be made.

But this conclusion assumes that the correlation between locus of control and verbal ability is unaffected by the differences in reliability. Unfortunately, the differences in reliability for the different levels of verbal ability can also lead to a spurious correlation between locus of control and ability. This occurs when the mean for all those who understand the items is definitely different from the mean achieved by responding randomly. The high ability children respond normally and show the usual mean. Those who cannot understand answer randomly and so receive the mean for random responding. The locus of control means for the low ability group therefore differ systematically from the locus of control means for the high ability group and cause the locus of control scores and verbal ability to correlate as an artifact of reliability. This spurious correlation is positive if the mean scores of those who understand the items is above the chance mean and negative if it is below that mean. Such a correlation can occur regardless of the actual underlying relationship between locus of control and verbal ability.

This appeared to be happening with both children's locus of control scales where they had a linear relationship between the group's ability level and its reliability. The locus of control mean of the children with reliable test scores was considerably above the chance level, whereas the low ability groups had means very close to the mean that would be generated by random responses. In both cases, there was also a significant

positive correlation between locus of control and verbal ability. This positive correlation may well be a function of the unreliability of the low ability children instead of being a function of a true relationship.

Naturally, if the locus of control scales have a spurious correlation with verbal ability, they will also have spurious correlations with all variables which normally correlate with verbal ability. This would be dangerous indeed.

It must also be remembered that mixing the locus of control items with other personality questionnaires, a common practice, reduced the reliabilities in a systematic but uninterpretable fashion. The Bialer and Miller scales appear to be sensitive to the context in which they are given, and more sensitive than instruments such as the children's social desirability scale. They are best given alone or mixed only with locus of evaluation items. Even then it would be good to check the internal consistency reliabilities to be assured that the scale is functioning properly.

It is therefore concluded that considerable caution must be exerted if an investigator is to use the Bialer or Miller scales with children. A better procedure may be to revise the scales for one's own situation as Epstein and Komorite (1971) did or turn to other methods of measuring the concept. Any new approach will need to be examined to see if the same problems arise as have been found here.

References

Bialer, I. Conceptionalization of success and failure in mentally retarded and normal children. Doctoral dissertation, Peabody College, 1960.

Bialer, I. Conceptionalization of success and failure in mentally retarded and normal children. Journal of Personality, 1961, 29, 303-320.

Blanton, R. C. & Nunnally, J. C. Evaluational language processes in the deaf. Psychological Reports, 1964, 15, 891-894.

Butterfield, E. C. Task recall and repetition choice as a function of locus of control components, mental age and skill vs. non-skill instruction. Unpublished doctoral dissertation, Peabody College, 1963.

Crandall, V. C., Crandall, V. J. & Katkovsky, W. A. A children's social desirability questionnaire. Journal of Consulting Psychology, 1965, 29, 27-36.

Cromwell, R. L.,Rosenthal, D., Shakow, D. & Zaber, T. P. Resistance to temptation, locus of control, choice behavior, and descriptions of parental behavior in schizophrenic and normal subjects. Journal of Personality, 1961, 29, 363-379.

Epstein, R. & Komoritz, S. S. Self esteem, success-failure, and locus of control in Negro children. Developmental Psychology, 1971, 4, 2-8.

Eysenck, S. B. G. Junior Eysenck Personality Inventory. San Diego, California: Educational and Industrial Testing Service, 1965.

James, W. H. Internal vs. external control of reinforcement as a basic variable in learning theory. Unpublished doctoral dissertation, Ohio State University, 1957.

Lefcourt, H. M. Internal vs. external control of reinforcement: a review. Psychological Bulletin, 1966, 65, 206-220.

Miller, J. O. Role perception and reinforcement conditions in discrimination learning among culturally deprived and non-deprived children. Unpublished doctoral dissertation, Peabody College, 1963.

Miller, M. B. Reliability of the Bialer-Cromwell scale in a mentally retarded group. Abstracts of Peabody Studies in Mental Retardation, 1960, 1, No. 73.

Nunnally, Jum C. Psychometric Theory. New York: McGraw-Hill, 1967.

Potter, R. B. & Cattell, R. B. Children's Personality Questionnaire. Champaign, Illinois: IPAT, 1963.

Rotter, Julian B. Generalized expectancies for internal vs. external control of reinforcement. Psychological Monographs, 1966, 80, No. 609, 28pp.

Rotter, J. B. Social learning and clinical psychology. Englewood Cliffs N. J.: Prentice-Hall, 1954.

Footnotes

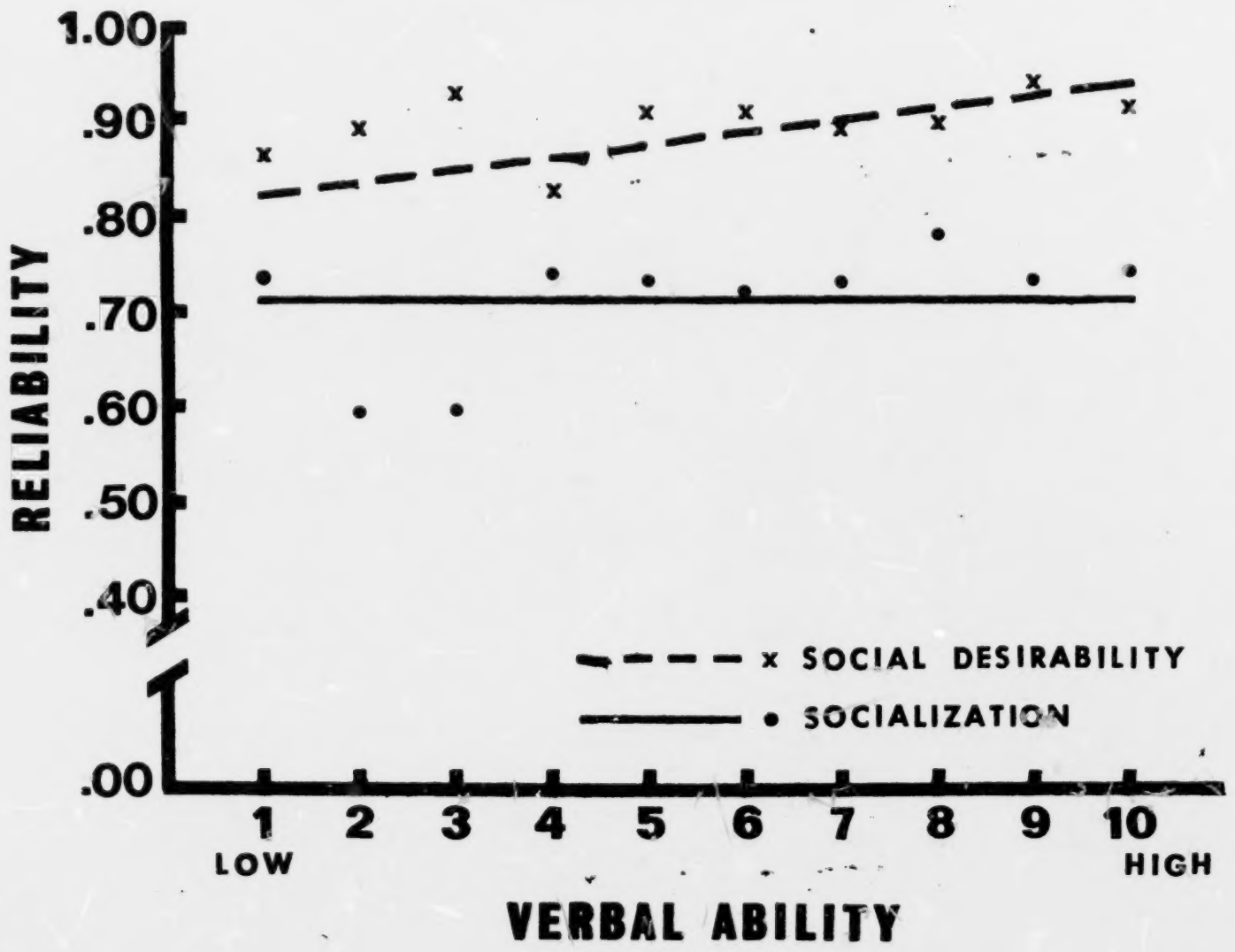
[1]This paper was supported in part by a grant (OEG-4-9-520427-0047) to R. L. Gorsuch from the U. S. Department of Health, Education, and Welfare, Office of Education, Bureau of Research.

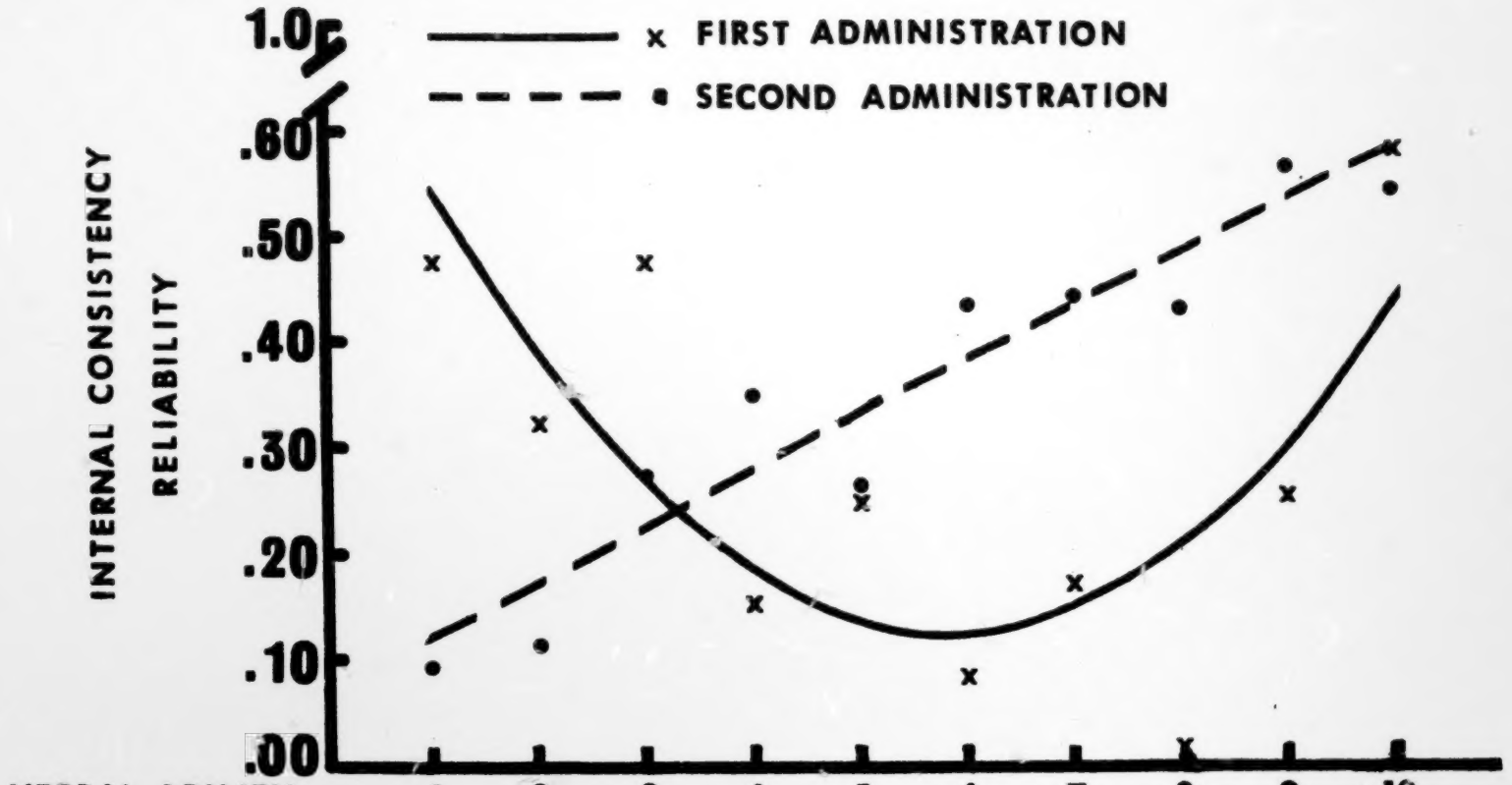
Figure Captions

Fig. 1. Reliabilities as a Function of Verbal Ability

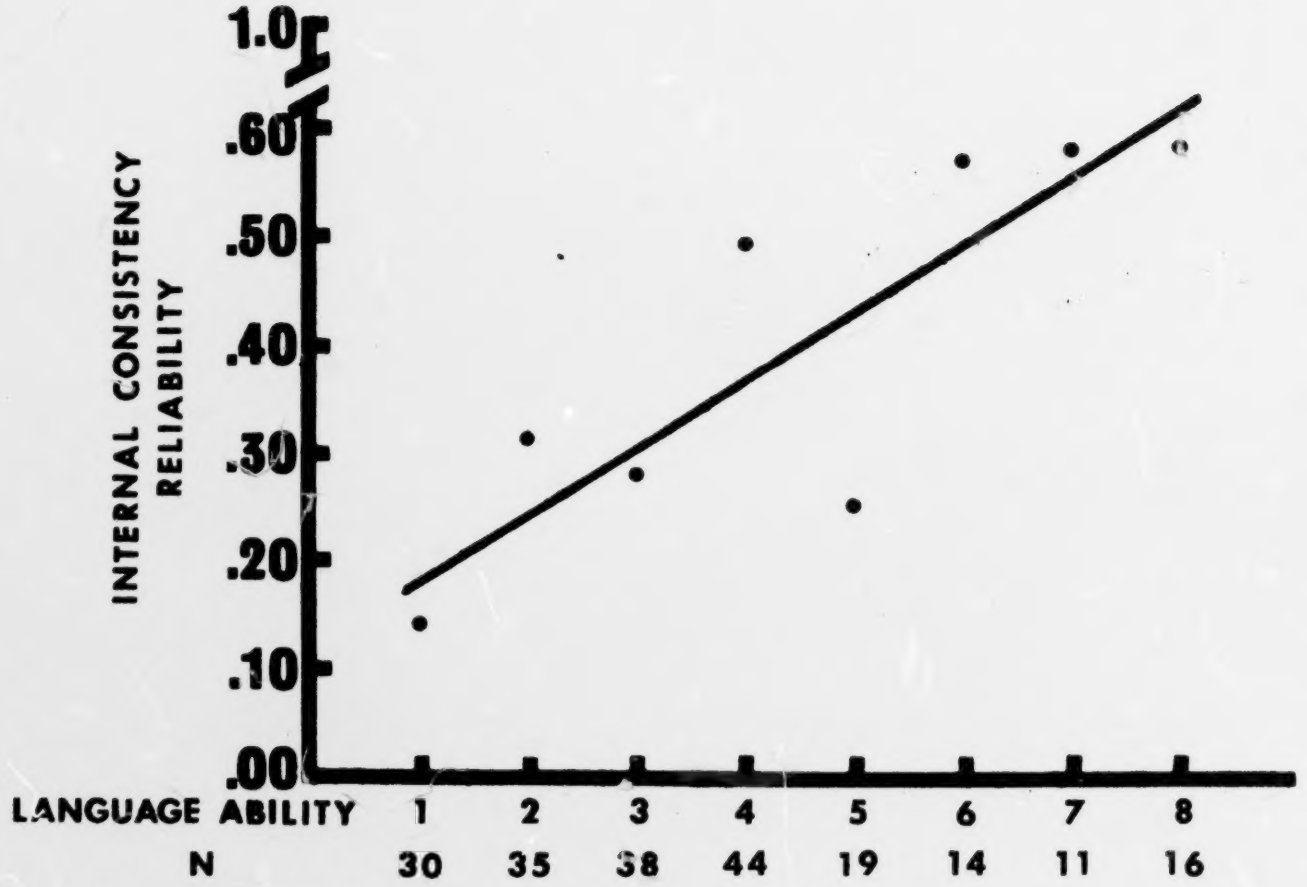
Fig. 2. Internal Consistency Reliabilities of Bialer's Children's Locus of Control Scale as a Function of Verbal Ability

Fig. 3. Internal Consistency Reliabilities of Miller's Children's Locus of Control Scale as a Function of Verbal Ability


1.00
.90
.80
.70
.60
.50
.40
.00
RELIABILITY
x SOCIAL DESIRABILITY
• SOCIALIZATION
1
2
3
4
5
6
7
8
9
10
LOW
HIGH
VERBAL ABILITY



VERBAL ABILITY	1	2	3	4	5	6	7	8	9	10
N: FIRST TESTING	16	23	29	48	55	80	50	57	42	20
N: SECOND TESTING	13	19	28	47	49	72	48	53	41	18


INTERNAL CONSISTENCY RELIABILITY
1.0
.60
.50
.40
.30
.20
.10
.00
LANGUAGE ABILITY 1 2 3 4 5 6 7 8
N 30 35 38 44 19 14 11 16

Appendix G

A Comparison of Several Techniques for Measuring Children's Values

A graduate student, Miss Joan Savarese, borrowed preliminary versions of some pilot value scales for use in her master's thesis, which was concerned with the effects of the race of the experimenter on value measurement in children. The project then borrowed her data to run analyses evaluating several possible approaches to value measurement with children.

Three measurement methods were used in the preliminary scales. The first followed Scott's approach (1965) which attempts to measure whether or not a person holds a value; the second was based on Sherif's work (e.g., Sherif, Sherif, & Nebergall, 1965) which gives value strength and ego involvement scores based on areas of acceptance and rejection; the third was a semantic differential (Osgood, et al, 1957) approach where value relevant activities are measured by evaluation and potency scales.

These three types of scales were included to 1) determine whether or not each approach gives adequate reliabilities when used with children of 4th, 5th and 6th grade (to our knowledge, several of these scales have not been previously used with this age); 2) examine the extent to which they measured the same aspect of values, and 3) enable preliminary item analyses to guide the development of the total item pool.

Subjects

232 4th, 5th, and 6th grade students in Murfreesboro school, each of whose parents had given permission for their child's participation, completed the questionnaires. For the purposes of the present analyses, the data was collapsed across experimenters so that race of experimenter was counterbalanced.

Method and Results

Internal Evidence on Functioning of the Scales

To determine whether or not the scales functioned as per expectations, internal consistency reliabilities were computed. A high internal consistency indicates that the children were responding systematically. This is a minimum criterion for the scale's functioning well for a particular age group.

The reliabilities for the Scott scales were corrected by the Spearman-Brown formula to approximate the length of scale used by Scott and the length which might be expected in the full study. These reliabilities ranged from .68 to .84, slightly below those of Scott. A slight lowering was expected since (1) the present work is with school children instead of college students, and (2) these scales had not gone through a scale development project as had Scott's.

The reliabilities of the five item semantic differential evaluation scales ranged from .76 to .94, indicating more than adequate functioning for research purposes. The potency scale with three items had lower reliabilities, .21 to .50, due to the inclusion of a poor item (hard-soft).

Meaningful internal consistency reliabilities cannot be computed for the Sherif scales.

The reliabilities indicated that these scales had the internal consistency that is necessary for use in our research.

Correlations Between the Scales

The values measured by some of the various scales overlapped. This allowed them to be correlated together to determine the extent to which they were alternative approaches to measuring the same values. Traditional attitude scales (e.g., Likert, Guttman and semantic differential) usually correlate quite high and are often interchangeable.

Correlations between Sherif scales and semantic differential scales measuring the same values ranged from .04 to .30. Correlations between Sherif scales and Scott scales measuring the same values were at the same level. On the other hand, correlations between Scott scales and semantic differential scales measuring the same values were between .14 and .39.

The various scales were parallel in content but generally contained different items. However, some of the items were identical in the Scott and Sherif scales. These were correlated to determine if the lack of scale correspondence was from the items shifting in the different formats or from a lack of similar content. Although items cannot be expected to correlate perfectly, these items correlated much lower than expected. The Scott vs. Sherif item correlations ranged from .16 to .26. (The semantic differential is so different that individual items were not compared). This suggests that the lack of overlap is from the way in which the task is defined rather than differences in content.

Item Analyses

Item analyses of the Scott and semantic differential scales were conducted. These results were used as a partial basis for the item development conducted within the grant itself.

Conclusions

1) The Scott scales appeared to be simple to understand and produced reliable measurements. If they had been as long as those used by Scott, the alpha reliabilities would have been almost the same. As a result, they will be used in further stages of the project.

2) The Sherif style scales seemed harder for the children. The children had to reread all items on their own after they were initially read to them, and then make their judgments. It's low relationships to the other two scales could be attributed to a lack of reliability or

measuring a different aspect of values. It was decided, therefore, to drop these scales from further stages of the project.

3) The semantic differential approach seemed to be easy to understand and had high internal reliability. However, the correlations with the other scales were not such as to assume that they were measuring values in the same way even though there was some minor overlap. It was decided, therefore, to keep this type of scale and to further investigate its relationship to Scott's approach.

References

Osgood, C. E., Suci, G. J. and Tannenbaum, P. H. The Measurement of Meaning. Urbana, Illinois: University of Illinois Press, 1957.

Scott, William. Values and Organizations. Chicago: Rand McNally, 1965.

Sherif, C. W., Sherif, M., and Nebergall, R. E. Attitude and Attitude Change: The Social Judgment-involvement Approach. Philadelphia: Saunders, 1965.

APPENDIX H

SEQUENTIAL MULTIPLE REGRESSION ANALYSES

Rationale

With only a few exceptions, psychological research proceeds by analyzing a linear model. A particular variable (called the predictor or independent variable) is used to estimate another variable (the criterion or dependent variable). In those cases where the hypotheses or data suggests curvilinear or interactive variables, derived or dummy variables are established to represent components of the curve or interaction. When properly developed, these derived variables are then related linearly to the dependent variable. Both analysis of variance and multiple regression analysis assume and analyze this single linear model (Cohen, 1968).

In analyses of how values related to value conflicts, grades, teacher perceptions of the child and the child's anxiety level, the hypotheses were multivariate and complex. Because analysis of variance has the restrictive assumptions that the predictors are uncorrelated, and the corollary of equal number of subjects per cell used to assure independence, analysis of variance procedures could not be run to test the hypotheses. Therefore, multiple regression analyses were undertaken.

In addition to allowing correlated predictors, regression analysis allowed several other advantages. First, regression analysis also provides the advantage of allowing continuous scores. Since most of the project variables were continuous, analysis of variance would require arbitrary splits of the variables. Allowing continuous variables to be included in the search for curvilinear relationships provides for the necessary knowledge to be developed by which appropriate splits can be made in later analyses if desired. Second, an overall test of significance was available wherever desired. In analysis of variance, 30 or 40 independent significance tests may be run but there is no overall test to see if the pattern of observed significance levels could have occurred by chance. It is simple to run such a test in a regression analysis. Third, this overall test could be computed for subsegments of the model. Do values in total add to the prediction? Such a question would be difficult to answer from analysis of variance approach except by examining several independent significance tests and attempting to generalize from these. A single overall significance test is available in regression analysis for this. Fourth, the hypotheses can be followed more closely in regression analysis since only those tests desired are actually run. This means that the investigator must think through what he is predicting -- a most helpful exercise -- and not simply hunt for post hoc explanations of that which is significant. Fifth, even though the interactions are significant in an analysis of variance, it generally tests main effects also. This is technically not appropriate but is generally done because of the way computer programs are constructed. Multiple regression analysis allows one to stop the analysis if an interaction is significant and thus avoid the technically inappropriate step. All these points

underscore the basic advantage of multiple regression for the project's purposes in addition to its ability to handle correlated models: it runs fewer significance tests and gives more overall tests of significance so that the probability of capitalizing upon a spurious chance significance level is notably less than when analysis of variance procedures are used.

Multiple regression analysis is not, however, without its problems. Not only does it require a greater degree of expertise to avoid errors in the analysis, it is costly. Analyses of variance are generally readily available in a pre-programed state at any computer center. Whereas running the regression analyses does not require extensive programing ability, it does require generating appropriate vectors to test the model in which one is interested. With the number of predictors involved in this study, and the fact that the work load goes up geometrically with the number of predictors, these regression analyses consummed an inappropriate amount of staff time and project funds. Doubling or tripling the number of subjects so that many could be randomly dropped to assure equal numbers of subjects per cell might have produced a higher cost/performance ratio.

Testing for Group Differences

The first step in Chapter VIII's analyses was to determine whether the predictors were functioning the same in the SES, sex, and race groups within the analysis. If this were not the case, then appropriate variables would be carried with the group interactions identified in the examination of the effects of values on the criterion and in developing the final model. This set of analyses was labeled for convenience sake the 100 Series.

Table 1 gives the procedure for the 100 Series of Analyses. Under each step, the full model of predictors is described as well as the restricted model. Following Bottenberg and Ward (1963), F-tests were computed by comparing the full model and the restricted model. If the same test is run by this procedure as is run in equivalent fixed model analysis of variance, the F ratio and degrees of freedom are identical. Also in the table is the decision rule established for each F-test. This indicates how that F-test affected the way in which the further analyses were done.

Step 101 in the 100 Series of Analyses is the overall test of significance. It assumes the interactions might aid in the prediction and includes all of the predictors. It simply tests whether or not the dependent measure can be significantly predicted at all. This is one of the tests which analysis of variance fails to compute.

Step 102 tests for the interactions with the groups. If these interactions were nonsignificant, then Step 103 tested for the main effects of SES, sex and race. If the interactions were significant, then further tests were run to determine the location of the significance.

Step 104 searched for non-value predictors which added to the significant interaction. If such a non-value predictor is found to interact, all of its group vectors were carried in further models and no further tests were run. That is, the test was not followed further to determine whether or not the significance was in the highest order interaction, in some lower order interaction or in the main effects. The reason further tests were not run was because the project was not that interested in the non-value measures. They were carried simply because it was felt that they would help to control for otherwise unpredictable variance in the criterion.

Steps 105, 106, 107, and 108 analyze the sources of significance for any value predictor when it was significant with its interactions. First, the SES x sex x race x value interaction was tested for significance. If this was significant, no further checks were run. If it were not significant, then three way, two way, and main effects significance tests were computed as appropriate.

When the 100 series levels of analyses was concluded, it was then reasonably apparent where the prediction was different for groups and where it was not.

Tests for Linear, Quadratic, and Cubic Value Predictions

Since some of the hypotheses suggested that the relationship between values and the criteria would be curvilinear, tests for these effects were run as the 200 Series of Analyses. When predicting anxiety and grades, the interactions between social desirability and the values were checked to determine whether or not response biases might be influencing the results. These interactions added nothing to the prediction in either case. This information, combined with previous information on the relationship of social desirability to the value scales, led to the decision that further checks on possible distortion of the value measures by social desirability were not worthwhile.

The 200 Series of Analysis is given in Table 2. The first test determined if any of the value predictors needed cubed terms in addition to raw and squared scores. If not, then a simpler-than-cubic model could be assumed. This test included all cubed terms at one time since no individual hypotheses regarding particular value measures had been developed. If the cubic prediction were significant, then an examination was made to see which value's cube scores added significantly to the prediction.

For those values which did not have cubic relationships to the criterion, Step 203 was run to test whether they had quadratic relationships. If not, then the simpler linear model was tested (Steps 205 and 206) .

This series of analyses indicated how the values predicted the criterion and tested whether each value added significantly to the prediction.

Developing the Final Model

In the sequential testing procedure, it was possible to make some of the tests somewhat conservative. This was done with the value measures. They had to prove that they could add to the prediction over and above the personality measures. This meant that any overlap between the values personality measures was attributed to the personality measures instead, even though some could argue it would be better attributed to the value measures.

The above analyses allowed conclusions to be reached as to the significant or non-significant contribution of each component of interest to this study. It was also of interest to provide a means by which the criterion or dependent variable can be predicted from the scores of the independent variables or predictors. For this reason, those variables which had all shown significant effects were included in one last model (cf. Table 3). Each element of this model was tested for significance since, for example, some of the interactions with the non-value predictors had not been further broken down. These tests for significance therefore allowed some conclusions to be reached as to which of the groups and variables were acting differently. They also were a check on the previous significance tests to assure that no insignificant effect had been erroneously included.

The model with only significant predictors was then re-computed. This model, along with the raw correlation coefficients, was used to construct the graphs and tables used to illustrate the significant effects.

Note that any relationship of the present procedure to stepwise multiple regression procedures is superficial. Stepwise procedures base their sequential testing only upon the data at hand. The present procedure based its sequential testing primarily upon the hypotheses and logic of the study. Whereas the stepwise procedure has many opportunities for capitalizing upon chance, the present procedure probably has as little or less opportunity of capitalizing upon chance than normal analysis of variance testing of hypotheses.

Table 1

100 Series of Analyses:
Does Adding Group Membership Effects Increase the Prediction Over and Above that from the Other Variables?

Step 101 - Test for Prediction of Dependent Variables

Full Model: 8 SES/Sex/Race subgroups plus these 8 subgroups in interaction with each of the predictors.

Restricted Model: No predictors.

Decision: If significant, go to Step 102. If not significant, STOP.

Step 102 - Test for Interactions with SES, Sex, and Race

Full Model: Same as Step 101.

Restricted Model: SES/Sex/Race vectors plus the rest of predictors but not interactions.

Decision: If interactions are significant, go to Step 104. If nonsignificant, go to Step 103.

Step 103 - Testing for Main Effects of SES, Sex and Race

Full Model: Predictors plus one vector for each of SES, sex and race.

Restricted Model: Predictors only.

Decision: If significant, keep group vectors and go to Step 201. If nonsignificant, drop group vectors and go to Step 201.

Step 104 - Search for Non-value Predictors Adding to the Significant Interaction

Full Model: Same as in Step 101.

For each predictor:

Restricted Model: Full model minus this predictor's interactions plus this predictor's original vector of scores.

Decision: If significant, keep this predictor's interactions. If nonsignificant, keep only the predictor's vector.

When all variables have been checked, go to Step 105.

Step 105 - Search for Value Predictors' Significant Interactions

Same as Step 104 except on values.

Table 1 (continued)

Decision: Drop nonsignificant interactions. Follow up significant interactions by going to Step 106. If no value interactions are significant, go to Step 201.

Step 106 - Testing Fourth Order Interactions with Value Predictions

Full Model: Predictors' interactions as appropriate from Steps 104 and 105. Otherwise, the predictors' vectors.

Restricted Model: Full model but with the value predictors' eight group interaction vectors replaced by the twelve three-way interaction vectors (note that these have 7 df).

Decision: If significant, go to Step 201. If nonsignificant, go to Step 107.

Step 107 - Testing Third Order Interactions

Full Model: Full model of Step 106 except that the vectors for the value predictor being examined are replaced as follows:

SES by Sex Interaction:
Full Model: Includes the four SES by Sex by value interaction vectors (with 4 df).
Restricted Model: Replace the four interaction vectors with vectors for the SES by value interaction and Sex by value interaction.

Repeat the above test for the SES by Race and Sex by Race interactions with the value predictor.

Decision: For any significant interaction, carry the interaction vectors into the model for Step 201. If one of the group variables enters into no significant interaction with the value predictor, go to Step 108. Otherwise, go to Step 201 if all the necessary value variable interactions have been analyzed.

Step 108 - Testing Group by Predictor Interactions

Full Model: Same as in 107 except that the value predictor's vectors would be formed by the group membership being analyzed; i.e., if sex is being analyzed, then male by value predictor and female by value predictor would be the vectors.

Restricted Model: Replace the interaction vectors with the value predictor vector.

Decision: If nonsignificant, drop from the model for Step 201.

Table 1 (continued)

Note: Repeat Steps 106-108 until all the value predictors' interactions are analyzed.

Table 2

200 Series Analyses: Do the Value Scores have Cubic, Quadratic, Linear or No Significant Relationships with the Dependent Variable Over and Above the Other Predictors?

Step 201 - Testing Whether any of the Value Predictions are Cubic rather than Quadratic/Linear

Full Model: Model resulting from prior steps plus the raw, squared and cubed scores for each value.

Restricted Model: Full model dropping all cubed scores.

Decision: If significant, go to Step 202. If nonsignificant, go to Step 203.

Step 202 - Testing Which of the Predictions are Cubic rather than Quadratic/Linear

Full Model: Same as Step 201.

For each value predictor:
Restricted Model: Full minus the value's cubed scores.
Decision: If nonsignificant, drop from model for next step. If significant, save in model for next step.

After testing all variables, go to Step 203.

Step 203 - Testing Whether any of the Predictions are Quadratic rather than Linear

Full Model: Appropriate model from prior steps plus each value and that value squared (i.e., full model of Step 201 minus the cubed scores). If coming from Step 202, carry the appropriate cubed scores.

For each value variable:
Restricted Model: Full model but without the squared scores for value where cubes were not significant.
Decision: If significant, keep quadratic for Step 204. If nonsignificant, drop quadratic and go to Step 205.

Step 204 - Testing Which of the Predictions are Quadratic rather than Linear

Full Model: Same as 203.

For each value predictor:
Restricted Model: Full minus that value's squared scores.
Decision: If nonsignificant, drop squared term from model for next step. If significant, save in model for next step.

Table 2 (continued)

After testing all variables, go to Step 205.

Step 205 - Testing Whether raw Value Scores add to the Prediction

Full Model: Model resulting from previous analyses with all linear, quadratic terms where quadratic was significant (same for significant cubic relationship).

Restricted Model: Drop all values without significant squared or cubed terms.

Decision: If significant, go to Stop 206. If nonsignificant, go to Step 301.

Step 206 - Testing each Value's Linear Prediction

Full Model: Same as in Step 205.

For each value predictor:
Restricted Model: Full minus that value's raw scores.
Decision: If significant, keep for model in Step 301.
If nonsignificant, drop from model.

Table 3

300 Series Analysis:
Determining Which Beta Weights are Significant

Step 301 - Testing Each Beta Weight

Full Model: Each non-value predictor in either vector or group membership interaction form. Values as appropriate from the 200 Series Analyses.

For each variable in full model:
Restricted Model: Full model minus this one vector.
Decision: Include in model if significant.